Power analysis for a proposed Group Randomized Control Trial (GRCT) on the Road to Mental Readiness (R2MP) program

Aihua Liu, Ph.D.
Biostatistician at Douglas Mental Health University Institute

Deniz Fikretoglu, Ph.D.
Defense Scientist at DRDC Toronto

Defence Research and Development Canada

Scientific Report
DRDC-RDDC-2014-R68
August 2014

Power analysis for a proposed Group Randomized Control Trial (GRCT) on the Road to Mental Readiness (R2MP) program

Aihua Liu, Ph.D.
Biostatistician at Douglas Mental Health University Institute

Deniz Fikretoglu, Ph.D.
Defense Scientist at DRDC Toronto

Defence Research and Development Canada

Scientific Report
DRDC-RDDC-2014-R68
August 2014

Abstract

The Road to Mental Readiness (R2MR) program is the largest mental health training initiative in the Canadian Armed Forces (CAF). As part of an effort to test the efficacy of R2MR at Basic Military Qualification (BMQ) with a group randomized control trial (GRCT), we conducted a robust power analysis to determine the sample size that would be required for the GRCT on R2MR. We also calculated intraclass correlation coefficients (ICCs) for the outcomes that will be measured in the GRCT, a necessary preliminary step for the power analysis. Data from the calculation of the ICCs were extracted from multiple programs of ongoing research with the Non-Commissioned Member (NCM) recruits, the intended target population for the GRCT. The results of our analyses suggest that data collected over the course of one full fiscal year will yield sufficient statistical power to detect expected effect sizes for most but not all of our outcomes. We therefore recommend data collection lasting up to one and a half years for the proposed GRCT on R2MR.

Significance to defence and security

This report provides the primary and secondary stakeholders for the program of research on R2MR (the Surgeon General and the Canadian Forces Leadership and Recruit School, respectively) with clear expectations about the duration of data collection for the proposed GRCT. This report also provides a model that other researchers can use to conduct power analyses in future, additional efficacy trials on R2MR.

Résumé

Le programme « En route vers la préparation mentale » (RVPM) est la plus grande initiative de formation en santé mentale des Forces armées canadiennes (FAC). Dans le cadre d'une démarche visant à mettre à l'épreuve l'efficacité de RVPM lors de la qualification militaire de base (QMB) au moyen d'un essai clinique randomisé (ECR) par grappes, nous avons mené une analyse d'efficacité rigoureuse afin de déterminer la taille de l'échantillon nécessaire à l'ECR par grappes de RVPM. Nous avons aussi calculé les coefficients de corrélation intraclasse (CCI) des résultats qui seront mesurés lors de l'ECR par grappes, étape préalable à l'analyse d'efficacité. Les données provenant du calcul des CCI ont été extraites de différents programmes de recherche continue portant sur les recrues militaires du rang (MR) , soit la population cible prévue de l'ECR par grappes. Les résultats de nos analyses laissent supposer que les données recueillies au cours d'une année financière entière offriront une puissance statistique suffisante pour cerner l'ampleur attendue de l'effet pour la plupart de nos résultats, mais pas pour tous. Nous recommandons une collecte de données pouvant durer jusqu'à un an et demi pour l'ECR par grappes proposé de RVPM.

Importance pour la défense et la sécurité

Le rapport présente les attentes claires des parties intéressées primaires et secondaires (le Médecin général et l'École de leadership et de recrues des Forces canadiennes, respectivement) du programme de recherche sur RVPM concernant la durée de la collecte de données de l'ECR par grappes proposé. Le rapport présente aussi un modèle que les autres chercheurs pourront utiliser pour mener des analyses d'efficacité lors d'essais d'efficacité futurs portant sur RVPM.

Table of contents

List of figures

List of tables

This page intentionally left blank.

1 Introduction

1.1 Background

There is increasing recognition in the Canadian Armed Forces (CAF) that maintaining good mental health is essential for optimizing force sustainability and operational effectiveness. The Road to Mental Readiness (R2MR) mental health education and training program was developed at the request of the Chief of Military personnel (CMP) and the CAF Surgeon General to help military members maintain good mental health throughout their career. R2MR is a large-scale mental health intervention with three key objectives:

- to increase mental health literacy (i.e., recognizing early signs and symptoms of mental health problems),
- to change negative attitudes towards mental health treatment, and
- to teach military members stress management skills they can use to maintain optimal mental health.

Importantly, an implicit assumption in R2MR is that a set of desired, short-term and long-term outcomes that are relevant in the military context will result from the uptake of these three key learning objectives. These outcomes include but are not limited to: increasing psychological resilience throughout the military career, decreasing psychological distress in the short term and decreasing the incidence and the severity of mental health problems in the long-term, increasing rates of help-seeking when mental health problems do arise, and ultimately, improving military training and operational performance outcomes both in the short- and long-term.

To achieve these short- and long-term objectives, R2MR is delivered throughout the military career cycle in the Army, and is being adopted in the other elements as well. Thus, various versions of R2MR exist: one designed specifically for Basic Military Qualification (BMQ) with Non-Commissioned Member (NCM) recruits, others specifically designed for primary and advanced leadership qualification (PLQ and ALQ, respectively), and others designed to be delivered specifically prior to and after an overseas deployment.

As a large-scale military mental health intervention, R2MR needs to be tested for efficacy in order to determine if (and to what extent) meaningful changes in the outcomes of interest are indeed taking place. While any of the existing R2MR versions could be tested for efficacy, a number of considerations favor choosing the BMQ version: first, the BMQ is military members' first exposure to R2MR and as such provides the foundation upon which all further mental health training is built. Therefore, ensuring that R2MR is efficacious at BMQ is critical for the success of all mental health training in the CAF. Second, BMQ is the only setting in which there is a captive audience/subject pool which makes an efficacy study feasible. And third, given the large number of NCM recruits who go through BMQ training on a continuous basis, the BMQ setting provides the largest sample size possible to detect what are likely to be small-size effects (1).

Randomized control trials (RCTs) are the gold standard for efficacy studies for a variety of interventions, including medical and/or mental health interventions such as R2MR. In the simplest type of RCT design, participants/individuals are randomly assigned to either an

intervention or a control condition. In settings where pre-existing clustering or grouping of individuals is present, where the intervention is delivered at the group (not the individual) level, and where there is "the risk of contamination"(2) —whereby group members randomized to the intervention condition could influence those randomized to the control condition through sharing the active ingredients of the intervention—it is more appropriate to randomize subjects at the group level, i.e., to conduct a groups randomized control trial (GRCT). In the case of the BMQ, individual recruits go through their 13-week training within a platoon (i.e., there is a pre-existing grouping or clustering of intervention targets), R2MR is delivered at the platoon (i.e., group) level, and the risk of contamination within a platoon (i.e., the group) cannot be ruled out. As such, testing the efficacy of R2MR requires a GRCT.

1.2 Methodological and statistical considerations in GRCTs

As stated in the previous section, in GRCTs, subjects are often linked through membership of a group. They have greater similarities within the group than individuals outside the group. Data collected from these groups are clustered, and we cannot assume statistical independence, i.e. subjects are not completely independent of each other. Consequently, compared to individual randomized trials where the statistical assumption of independence within the sample is warranted, group randomized trials have less information contributed by each individual. This results in reduced statistical power for detecting significant intervention effects when conducting analyses at the individual level. In the extreme case where all the individuals in a group have the same outcome, (i.e., where group members are completely dependent), the sample size contribution from the group is 1 rather than the number of individuals in the group. Thus, power and sample size calculation for group randomized trials has to take into consideration the within-group clustering effect. Intra-class correlation coefficient (ICC) is the most often used measure of this effect in group randomized control trials.

1.3 Intra-class Correlation Coefficient (ICC)

In a group randomized control trial, the total variability of an outcome is comprised of two parts: the within-group variation and between-group variation. ICC measures the proportion of the total variance of as:

$$\rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2} \tag{1}$$

where, σ_b is the between-group variance and σ_w is the within-group variance.

The value of ICC can range from 0 to 1. A value of 0 means that all the variance of the outcome is due to the within group variation and there is no between group variation, i.e., the individuals within a group are completely independent. In this case, the group randomized trial can be treated as an individual randomized trial for power and sample size calculation. In the complete opposite scenario, where individuals within a group are completely dependent, the between group variation is the only source of variance in the outcome. In this situation, ICC is 1, and the power for detecting significant intervention effects is greatly reduced.

2 Objective

The overall objective of this report is to conduct a robust power analysis to determine the sample size that would be required for the GRCT on R2MR. An intermediate step required for the power analysis is the calculation of the ICCs for the outcomes that will be measured in the GRCT on R2MR.

Conducting and reporting a detailed power analysis is one of the critical recommendations in the Consolidated Standards of Reporting Trials (CONSORT) for cluster (group) randomized control trials (3). Conducting a power analysis also sets reasonable expectations around what sample size may be required to detect various intervention effects. Given that R2MR at BMQ will require data collection in an operational/training setting [i.e., the Canadian Forces Leadership and Recruit School (CFLRS)], school administrators will want to know at what point a large efficacy trial on R2MR may be reasonably expected to end. A power analysis provides reasonable expectations for the length of that data collection. A robust power analysis also guards against under- and over-recruitment of subjects; "studies are not just wasteful when they stop too early [i.e., under-recruitment], they are also wasteful when they stop too late [i.e., over-recruitment]" (4). Furthermore, both scenarios are considered unethical by having exposed subjects to unnecessary risk under the principles of the World Medical Association Declaration of Helsinki (5).

A separate report on the results of the power analysis was requested by the primary and secondary stakeholders of the R2MR program of research so that the report could be used to inform discussions and expectations around subject recruitment among the stakeholders and the research team prior to the beginning of the efficacy trial.

The authors also see value in publishing the results of this power analysis as a separate report for the larger defence scientific community for the following reasons: First and foremost, while most data in a military setting is clustered in nature (clustered within units, such as platoons, brigades, battalions, regiments), few researchers are familiar with the tools to determine the extent of clustering (i.e., by way of calculating ICCs), the implications of clustering for whether or not some of the assumptions of commonly used statistical tests are violated, and alternative methods for conducting power analysis and common statistical analyses while taking into consideration the clustered nature of the data. This report provides a model that can be used to determine the extent of clustering in research data, outlines the implications of clustering for data analysis and power analysis, and shows how clustering can be taken into account in conducting a power analysis.

3 Methods

3.1 Selection of outcome measures and data extraction for the calculation of ICCs

Psychological Outcomes: We calculated ICCs for two psychological outcomes: The Patient Health Questionnaire – 9 (PHQ-9 (6)), a 9-item self-report measure of depression in the last two weeks that assesses depressive symptoms based on the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (7), and the Patient Health Questionnaire -15, (PHQ-15 (8)), a 15-item self-report measure which assesses somatic symptoms in the last two weeks. The reliability, validity, sensitivity, and specificity of these two measures are well-established in extant literature (6, 9, 10). These two measures were selected as they are similar to psychological outcome measures that will likely be used in the GRCT and also because data are routinely collected on these two measures in the first few weeks of recruit training at CFLRS part of an ongoing health surveillance project, (i.e., the Recruit Health Questionnaire Study). We extracted data from N=3301 recruits in the RHQ database (reflecting n=75 platoons); ICC calculations were performed on anonymized data.

Performance Outcomes: We calculated ICCs for one performance outcome that we will also use in the GRCT – graduation from the BMQ. This is a binary outcome (pass/fail) that is routinely collected administratively by CFLRS. Data from three recent fiscal years (2010-11, 2011-12, and 2012-13) was obtained through the Commanding Officer (CO) at CFLRS for the calculation of the ICCs (11). The anonymized data extracted (including N= 7501 recruits and 227 platoons)) were used for the calculation of ICCs.

Mental Health Treatment Attitude Outcomes: We calculated ICCs for eight constructs related to attitudes towards seeking mental health treatment: Overall attitudes, instrumental attitudes, affective attitudes, overall intention, overall perceived norms, overall perceived control, perceived control over seeking treatment, and perceived self-efficacy for seeking treatment. These constructs were assessed with the Canadian Armed Forces Recruit Mental Health Service Use Questionnaire (CAF-MHSUQ) (12, 13); a measure designed specifically for the target GRCT population. The internal consistency and factorial validity of this new measure has been established in a series of studies (12, 13). Data were extracted from a study examining the uptake of R2MR concepts under various conditions (14); ICCs were calculated on approximately N=308 recruits and N=6 platoons.

3.2 ICC calculation

ICCs are ideally estimated using pilot data (15). In the absence of pilot data, estimates are based on what has been reported in existing literature on similar interventions with similar target populations. Three estimation methods are commonly used for calculating ICCs in group randomized trials: analysis of variance, mixed effects models, and generalized estimating equations. For continuous outcome variables, linear mixed effects model (in which the group is treated as a random effect) is the most popular approach for ICC estimation. This approach has the advantage of calculating the values of within-group and between-group variance. Another

important advantage of using linear mixed effects models is that this method avoids having negative estimates for ICC. Although it is considered impossible for true ICC values to be negative, when using analysis of variance or generalized estimating equations, negative ICC values can sometimes arise. The negative ICC values are believed to be due to chance and are often truncated to 0 (16). Given all its advantages, we employed linear mixed effects models for estimating ICC for continuous outcomes in this project. For binary outcome variables, estimating ICC is much more complicated. The overall value of ICC is affected by the prevalence of the outcome, and whether success rates are the same for the intervention and the control groups (16, 17). Among the several approaches that can be used for estimating ICC for binary outcome variables, the generalized estimating equation methods, which provides more accurate overall ICC estimates especially when the success rates are not similar between the two groups, is recommended (17).

3.3 Overall analytic method for the power analysis

We employed two approaches for the power analysis presented in the current report. First, we calculated power for each outcome given the expected sample size and the desired intervention effect. Second, we estimated the minimum detectable intervention effect based on the expected sample size and desired power. These two approaches capture the range of conditions under which the proposed GRCT on R2MR will be able to optimally test the effects of R2MR as an intervention.

We calculated the expected sample size based on administrative CFLRS data described in a DRDC Toronto Technical Memo (11). This document indicates that the average platoon size at intake at Basic Military Qualification (BMQ) ranges from 50 to 60. During one fiscal year, about 40-50 platoons go through BMQ training and are available for participation in the GRCT. In previous research with NCM recruits at CFLRS, participation rates varied from 50% - 70% across different platoons. Based on these numbers, we created four scenarios for power analyses: assuming, 1) the lowest participation rate (=50%) or the highest participation rate (=70%); and 2) the lowest number of recruited platoons (=40) or the highest number of recruited platoons (=50). We used 55 as the average platoon size. Based on this calculation, the expected sample size for the GRCT during a full fiscal year ranges approximately from 1100(=55*50%*40), calculated from the worst case scenario where the lowest participation rate and lowest number of recruited platoons are assumed, to 1925 (=55*70%*50), calculated from the best case scenario where the highest participation rate and the greatest number of recruited platoons are assumed. Naturally occurring dropouts from BMQ (through release or attrition) were taken into account in the power analysis for BMQ graduation rate. Based on the same DRDC Toronto Technical Memo (11), the BMQ dropout rate is around 15% (12% - 19% in the last three fiscal years). Thus, in the power analysis for BMQ graduation rate, the average platoon size is 47, which reflects a reduction of 15% (=55*0.85).

For our power analysis, ICCs for continuous outcomes are estimated using data extracted from previous studies with the GRCT study population of NCM recruits. Desired power is set as 80%. The intervention effect, quantified using the upper limit of the effect sizes for continuous outcomes reported in previous military mental health interventions (18, 19), is set as 0.2. An effect size of 0.2 means that we expect to detect intervention effects that will make the intervention (R2MR) group differ from the control (no R2MR) group by at least 0.2 units of the

population standard deviation of the outcome. For the binary outcome BMQ graduation rate, the intervention effect is quantified as the increase in the success rate from the control group to the intervention group. The BMQ graduation rate is approximately 80% based on historical CFLRS administrative data; Given that these administrative data predate the introductions of the current version of R2MR, the 80% graduation rate is assumed to be what we might expect to see in the control group of the GRCT. We assume that the intervention may increase the graduation rate by 10% which renders a graduation rate of 90% for the intervention group. Estimates for the proportion of variance explained by group-level covariates were not available from any previous pilot study in NCM recruits, and were therefore determined based on recommendations in existing literature (20), a common approach in the absence of data from the target study population.

We used the Optimal Design Plus Empirical Evidence version 3.0 software (21) for all power analyses; this software is designed for conducting power and sample size analyses for detecting significant differences between the intervention and control groups specifically in GRCTs.

4 Results

Table 1 shows the ICC values calculated for each of the continuous outcomes (2nd column). These values suggest that there is a clustering effect for some outcomes. For example, for the Self-efficacy for seeking mental health treatment score, the ICC is 0.038, indicating that the within-cluster variation accounts for 3.8% of the total variance for this variable. Similarly, the ICC calculated for the binary outcome BMQ graduation rate (=0.020, shown in the 2nd column of Table 2), also suggests the existence of a clustering effect. Other variables show a clustering effect to varying degrees as well; these are summarized in Tables 1 and 2.

Table 1: *Power calculation for detecting significant intervention effects with a desired effect size=0.2 for continuous outcomes under four different scenarios for expected sample size.*

Outcome	ICC	Power			
		Number of recruited platoons=40		Number of recruited platoons=50	
		Participation rate=50%	Participation rate=70%	Participation rate=50%	Participation rate=70%
Overall Attitude	0.015	> 80%	> 90%	> 90%	> 95%
Instrumental Attitude	0.023	80%	> 85%	> 85%	> 90%
Affective Attitude	0.006	> 85%	> 95%	> 90%	> 95%
Overall Intention	0.008	> 85%	> 90%	> 90%	> 95%
Overall Perceived Norms	0.009	> 85%	> 90%	> 90%	> 95%
Overall perceived Control	0.027	> 75%	> 85%	> 85%	> 90%
Perceived control	0	> 90%	> 95%	> 95%	> 95%
Self-efficacy	0.038	> 70%	> 80%	> 80%	> 85%
PHQ9	0.007	> 85%	95%	> 90%	> 95%
PHQ15	0.025	> 75%	> 85%	> 85%	> 90%

Other parameters used for calculating the power: Significance level α =0.05, number of subjects in each platoon at intake=55, proportion of variance explained by group level covariates: 0.4.

Software used for power calculation: Optimal Design Plus Empirical Evidence version 3.0 (21).

Table 1 summarizes the estimated statistical power for continuous outcomes under the four different scenarios for expected sample size. It can be seen that in three out of the four scenarios, for all of the outcomes there is very good (>90%) or sufficient power (>80%) for detecting a significant intervention effect of 0.2 or higher. Even in the worst case scenario where the lowest number of platoons are recruited (=40) and the lowest participation rate (=50%) is achieved, there will be sufficient power for 7 out of the 10 outcomes.

For the binary outcome BMQ graduation, Table 2 shows the estimated statistical power for detecting the expected difference in success rate between the intervention and the control groups. It can be seen that in all of the four scenarios, there will be excellent power (> 95%) for detecting an intervention effect that produces a 10% increase in the success rate.

Table 2: *Power calculation for detecting significant intervention effects for the binary outcome BMQ graduation rate under four different scenarios for expected sample size.*

Outcome	ICC	Power			
		Number of recruited platoons=40		Number of recruited platoons=50	
		Participation rate=50%	Participation rate=70%	Participation rate=50%	Participation rate=70%
BMQ Graduation	0.020	> 95%	> 95%	> 95%	> 95%

Other parameters used for calculating the power: Significance level α =0.05, number of subjects in each platoon at intake=55, BMQ graduation rates in the control and intervention groups are 80% and 90%, respectively.

Software used for power calculation: Optimal Design Plus Empirical Evidence version 3.0 (21).

Table 3 summarizes the results from the minimum detectable effect size calculation for the continuous outcomes. It shows that in all four scenarios, for all outcome variables, there is sufficient power for detecting intervention effects, with effect size as small as 0.22. This value of minimum detectable size is improved to be 0.20 when excluding the worst scenario. For some outcomes, there is sufficient power for detecting intervention effects with even smaller effect sizes. For example, for PHQ-9 depression symptom scores, the minimum detectable effect size is 0.18 in the worst case scenario, indicating that an intervention effect that increases the mean value of the depression symptom scores by 0.18 unit of population standard deviation could be detected as statistically significant, even if subject recruitment ends up yielding the smallest expected sample size.

Table 3: *Minimum detectable effect size calculation for continuous outcomes under four different scenarios for expected sample size.*

Outcome	ICC	Minimum detectable effect size			
		Number of recruited platoons=40		Number of recruited platoons=50	
		Participation rate=50%	Participation rate=70%	Participation rate=50%	Participation rate=70%
Overall Attitude	0.015	0.19	0.17	0.17	0.15
Instrumental Attitude	0.023	0.20	0.18	0.18	0.16
Affective Attitude	0.006	0.18	0.16	0.16	0.14
Overall Intention	0.008	0.19	0.16	0.17	0.14
Overall Perceived Norms	0.009	0.19	0.16	0.17	0.15
Overall perceived Control	0.027	0.21	0.19	0.19	0.17
Perceived control	0	0.17	0.15	0.16	0.13
Self-efficacy	0.038	0.22	0.20	0.20	0.18
PHQ9	0.007	0.18	0.16	0.16	0.14
PHQ15	0.025	0.21	0.19	0.18	0.17

Other parameters used for calculating the minimum detectable effect size: power=0.8, significance level α =0.05, number of subjects in each platoon at intake=55, proportion of variance explained by group level covariates: 0.4.

Software used for power calculation: Optimal Design Plus Empirical Evidence version 3.0 (21).

Table 4 shows the minimum detectable intervention effects for the binary outcome BMQ graduation rate. The 2nd column shows the success rate for the control group which, as stated previously, was obtained based on CFLRS administrative data (11). For each of the four scenarios, we calculated the minimum detectable success rates in the intervention group that are statistically different from that in the control group (shown in the 3rd column). The results indicate that subject recruitment over one full fiscal year will provide us sufficient power to detect a success rate in the intervention group as low as 86% - 88%, meaning a 6-8% increase in the BMQ graduation rate produced by the intervention.

Detailed results for estimated power, minimum detectable effect size for continuous outcomes, minimum detectable success rates for the binary outcome are presented in appendix.

Table 4: *Minimum detectable success rate in the intervention group for binary outcome BMQ graduation rate under four different scenarios for expected sample size.*

Outcome	Estimated success rate in the control group	Minimum detectable success rate in the intervention group			
		Number of recruited platoons=40		Number of recruited platoons=50	
		Participation rate=50%	Participation rate=70%	Participation rate=50%	Participation rate=70%
BMQ Graduation	80%	88%	87%	87%	86%

Other parameter used for calculating the minimum detectable success rate among the intervention group: power=0.8, significance level α =0.05, number of subjects in each platoon at intake=55,

Software used power calculation: Optimal Design Plus Empirical Evidence version 3.0 (21).

5 Discussion

The overall objective of this report was to conduct a robust power analysis to determine the sample size that would be required for the GRCT on R2MR. As a preliminary step towards conducting a power analysis, we also calculated ICCs for the outcomes that will be measured in the GRCT on R2MR.

The results of our work suggest that for the proposed GRCT on R2MR, we can expect that data collected over the course of one full fiscal year will yield sufficient statistical power to detect the expected effect sizes for most of our outcomes. Assuming a start date of fall 2014 for the GRCT, we expect that data collection at CFLRS would end around fall to winter 2015. Given that data will be monitored throughout the GRCT and analyzed at various intervals, we expect a final analysis and report to be complete by early spring 2015. These reports will be disseminated among the primary and secondary stakeholders for this project (the Surgeon General and the CO at CFLRS) in face-to-face meetings.

While we made every effort to identify variables that are close to the ones that will be used in the GRCT as outcomes (and for which pilot data exist), there are a number of outcome variables, such a psychological resilience, where we could not locate existing pilot data for our study population. A number of authors have argued for an upper limit of effect sizes of 0.2 and ICCs of 0.05 (1) for military mental health outcomes in GRCTs, and it is possible to use these estimates to arrive at the sample sizes that will be required to detect intervention effects for variables for which data do not exist. A scenario based on those upper limit estimates closely mirrors that for the self-efficacy for mental health treatment variable in Table 1.

We also note here that the naturally occurring dropouts from BMQ (through release or attrition) were taken into account in power analysis only for the binary outcome of BMQ graduation rate but not for other outcomes. As outlined in the DRDC Toronto Technical Memo which summarized administrative CFLRS data from three recent fiscal years (11), the dropouts in BMQ training tend to occur at different stages in the 13-weeks of BMQ training. In the GRCT, since all the outcomes except BMQ graduation rate will be evaluated at more than one time point through the 13 weeks, we expect that for all of the recruits at intake, we will have data for these outcomes from at least one time point. This will allow us to retain these recruits in future statistical analyses since mixed models analysis – which has the ability to accommodate missing date points (22, 23), will be employed for modeling these outcomes.

In summary, using existing pilot data from administrative datasets and large studies conducted in our target GRCT population, taking into account dropouts, and considering possible scenarios for variables for which we do not have pilot data, the power analysis presented in this report suggests that data collection over a full fiscal year should be sufficient for most of our outcomes of interest, with the caveat that for some of our outcomes, it may be necessary to stretch the data collection by 3-4 months.

In this report, the ICCs calculated for the outcome variables of interest ranged from 0 to 0.038. Thus, our ICCs are quite small. ICCs of 0.05, 0.10, and 0.15 are considered small, medium, and large, respectively (24). However, as has been noted in the literature, the magnitude of the clustering effect depends not just on the magnitude of ICCs but also the size of the clusters. Even

small ICCs, when accompanied by large cluster sizes (as is the case in this report with typical platoon sizes of about 55), can lead to significant reductions in statistical power (25) and "can still affect the validity of conventional statistical analyses" (pp- 199-200). We therefore caution researchers not to dismiss small ICCs without first carefully considering the cluster size (and the overall clustering effect).

In addition to calculating ICCs and not dismissing out of hand small ICCs, researchers must also determine whether their primary objective is to control for (which is the case in our planned GRCT) the existing clustering effects or to discover the clustering groups. Different analytic strategies exist for these two objectives; discussion of these various analytic strategies are beyond the scope of this report but can be found in literature (26-28).

Furthermore, the issue of when to suspect relatively small versus relatively large clustering effects must be carefully considered. In the recruit training context, in which relative strangers come together to form a platoon, we expect to find small clustering effects at the beginning of training; this is indeed what we find in the current report where most of the data come from the first few weeks of recruit training. However, it is entirely possible that these effects may be larger as individuals spend more and more time together as a cluster/unit within their platoon over the 13-week recruit training. The same logic applies to research conducted with populations other than military recruits, where the units and clusters have been in existence longer; here, we may expect the clustering effect to be larger. Such research scenarios will call for careful consideration of analytic methods that take into consideration the clustered nature of the resultant data.

References.

[1] Bliese PD, Adler AB, Castro CA. Research-Based Preventive Mental Health Care Strategies in the Military. In: Adler AB, Bliese PD, Castro CA, editors. Deployment psychology : evidence-based strategies to promote mental health in the military. 1st ed. Washington, DC: American Psychological Association; 2011. p. xi, 294 p.

[2] Christie J, O'Halloran P, Stevenson M. Planning a cluster randomized controlled trial: methodological issues. Nursing research. 2009;58(2):128-34.

[3] Campbell MK, Piaggio G, Elbourne DR, Altman DG, Group C. Consort 2010 statement: extension to cluster randomised trials. Bmj. 2012;345:e5661.

[4] Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, et al. Power failure: why small sample size undermines the reliability of neuroscience. Nature reviews Neuroscience. 2013;14(5):365-76. Epub 2013/04/11.

[5] Helsinki WMAdo. Recommendations guiding physicians in biomedical research involving human subjects. JAMA: the journal of the American Medical Association. 1997;277(11):925-6.

[6] Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA : the journal of the American Medical Association. 1999;282(18):1737-44.

[7] Association AP. Diagnostic and statistical manual of mental disorders (4th ed.). Washington, DC: Author1994.

[8] Kroenke K, Spitzer RL, Williams JB. The PHQ-15: validity of a new measure for evaluating the severity of somatic symptoms. Psychosomatic medicine. 2002;64(2):258-66.

[9] Spitzer RL, Williams JB, Kroenke K, Hornyak R, McMurray J. Validity and utility of the PRIME-MD patient health questionnaire in assessment of 3000 obstetric-gynecologic patients: the PRIME-MD Patient Health Questionnaire Obstetrics-Gynecology Study. American journal of obstetrics and gynecology. 2000;183(3):759-69.

[10] Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. Journal of general internal medicine. 2001;16(9):606-13.

[11] Richards K, Fikretoglu D. Using Administrative Data to Inform the Design of a Group Randomized Control Trial of the Efficacy of the Road to Mental Readiness Program. Defence R&D Canada, Toronto research centre, 2013.

[12] Fikretoglu D, Blais A-R, Lam Q, Sullivan-Kwantes W, Richards K, McCreary D. Changing mental health care seeking attitudes to build psychological resilience - Lessons learned from a Canadian military intervention. The International Society of Traumatic Stress Studies (ISTSS); November 6-9; Philadelphia, PA, USA2013.

[13] Blais A-R, Fikretoglu D, Lam Q. Refinement and translation of an instrument to assess mental healthcare attitudes at the beginning of the military career. Military Veteran Health Research Forum (MVHRF); November 25-27; Kingston, ON2013.

[14] Fikretoglu D, Beatty E. Comparing different versions of Road to Mental Readiness to determine optimal content. DRDC Toronto Ethics Protocol, 2013.

[15] Killip S, Mahfoud Z, Pearce K. What is an intracluster correlation coefficient? Crucial concepts for primary care researchers. Annals of family medicine. 2004;2(3):204-8.

[16] Eldridge S, Kerry S. A practical guide to cluster randomized trials in health services research: John Wile & Sons, Ltd; 2012.

[17] Wu S, Crespi CM, Wong WK. Comparison of methods for estimating the intraclass correlation coefficient for binary responses in cancer prevention cluster randomized trials. Contemp Clin Trials. 2012;33(5):869-80. Epub 2012/05/26.

[18] Adler AB, Bliese PD, McGurk D, Hoge CW, Castro CA. Battlemind debriefing and battlemind training as early interventions with soldiers returning from iraq: Randomization by platoon. Journal of consulting and clinical psychology. 2009;77(5):928-40.

[19] Castro CA, Adler AB, McGurk D, Bliese PD. Mental health training with soldiers four months after returning from Iraq: randomization by platoon. Journal of traumatic stress. 2012;25(4):376-83.

[20] Bloom HS, Richburg-Hayes L, Black AR. Using Covariates to Improve Precision: Empirical Guidance for Studies that Randomize Schools to Measure the Impacts of Educational Interventions. Educational Evaluation and Policy Analysis. 2007;29(1):30-59.

[21] Spybrook J, Bloom H, Congdon R, Hill C, Martinez A, Raudenbush S, inventors; Optimal Design Plus Version 3.02011.

[22] Feng Z, McLerran D, Grizzle J. A comparison of statistical methods for clustered data analysis with Gaussian error. Stat Med. 1996;15(16):1793-806. Epub 1996/08/30.

[23] Keselman HJ, Algina J, Kowalchuk RK. The analysis of repeated measures designs: a review. Br J Math Stat Psychol. 2001;54(Pt 1):1-20. Epub 2001/06/08.

[24] Hox JJ. Multilevel analysis : techniques and applications. Mahwah, N.J.: Lawrence Erlbaum Associates; 2002. x, 304 p. p.

[25] Zyzanski SJ, Flocke SA, Dickinson LM. On the nature and analysis of clustered data. Annals of family medicine. 2004;2(3):199-200. Epub 2004/06/24.

[26] Donner A, Klar N. Design and analysis of cluster randomization trials in health research. London New York, N.Y.: Arnold; Co-published by the Oxford University Press; 2000. x, 178 p. p.

[27] Cattell RB. The description of personality: Basic traits resolved into clusters. Journal of Abnormal and Social Psychology. 1943;38:476-506.

[28] Tryfos P. Methods for business analysis and forecasting : text and cases. New York: Wiley; 1998. xiv, 576 p. p.

This page intentionally left blank.

Annex A Figures for power and minimum detectable effect size calculation

Figure A.1: *Power calculation for continuous outcome variable overall attitude.*

From empirical analyses: ICC = 0.015

Assuming effect size is 0.2, proportion of the variance explained by group level covariates: 0.4. Number of subjects in each platoon is 28 (participation rate=50%, Figure A.1.a) to 39 (participation rate=70%, Figure A.1.b). Significant level: a=0.05.

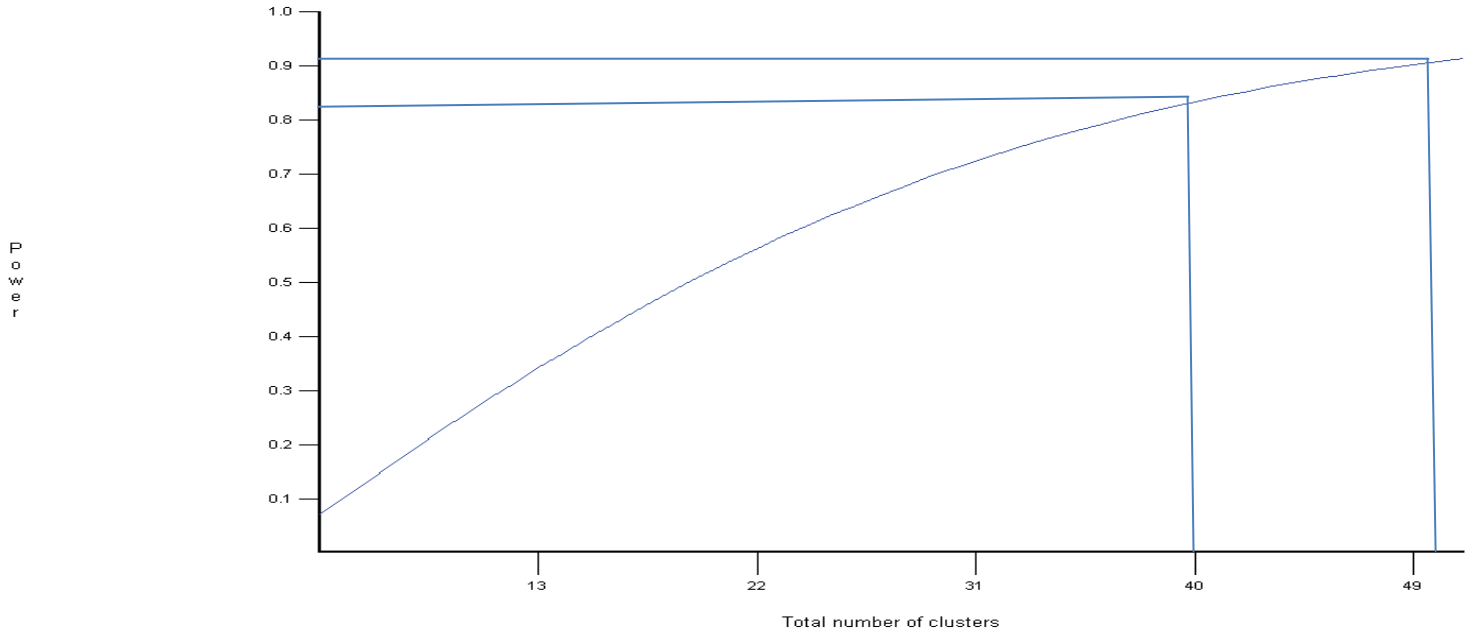

Figure A.1.a

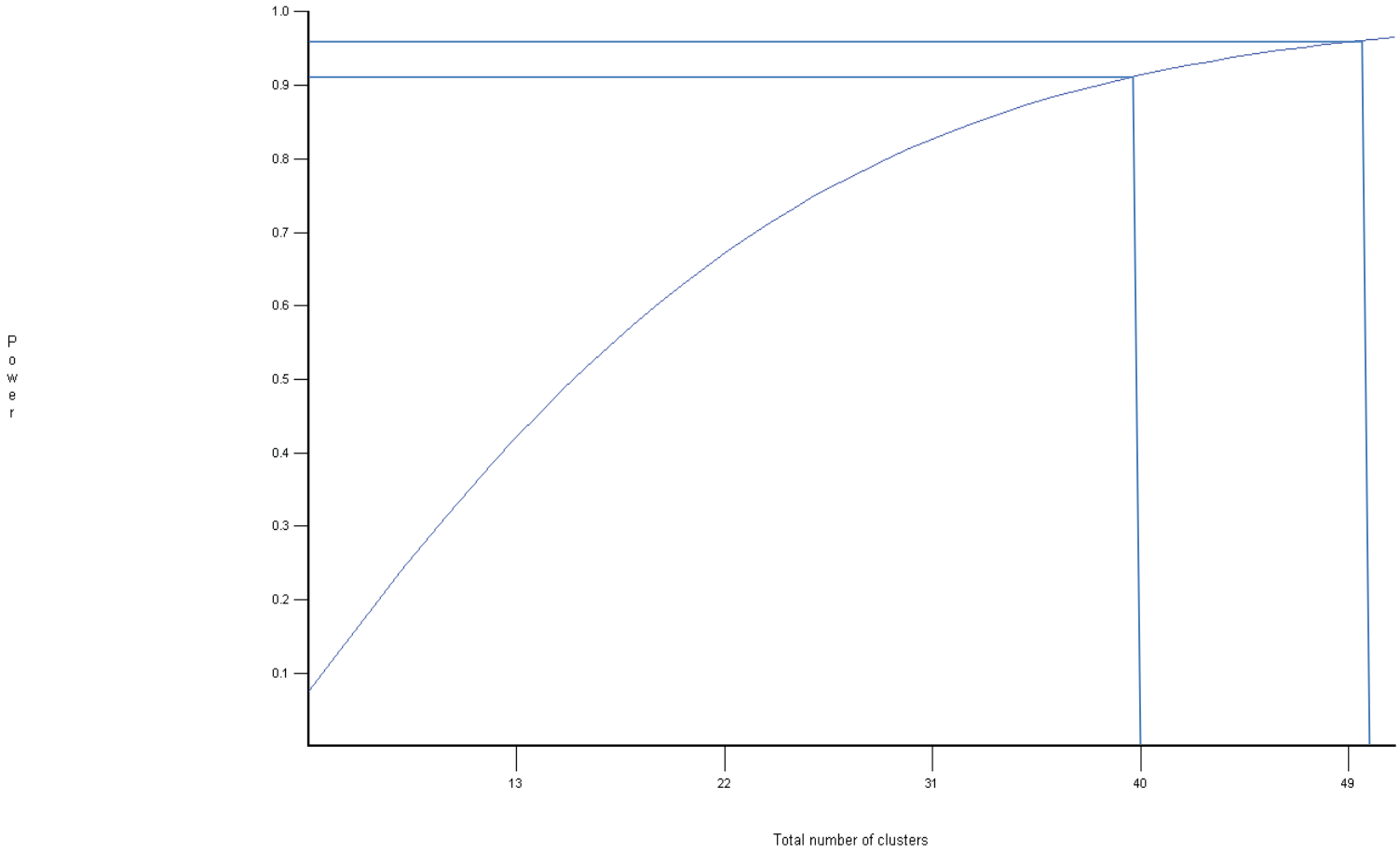
Power
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
13
22
31
40
49
Total number of clusters

Figure A.1.b

***Figure A.2**: Power calculation for continuous outcome variable instrumental attitude*

From empirical analyses: ICC = 0.023

Assuming effect size is 0.2, proportion of the variance explained by group level covariates: 0.4. Number of subjects in each platoon is 28 (participation rate=50%, Figure A.2.a) to 39 (participation rate=70%, Figure A.2.b). Significant level: a=0.05.

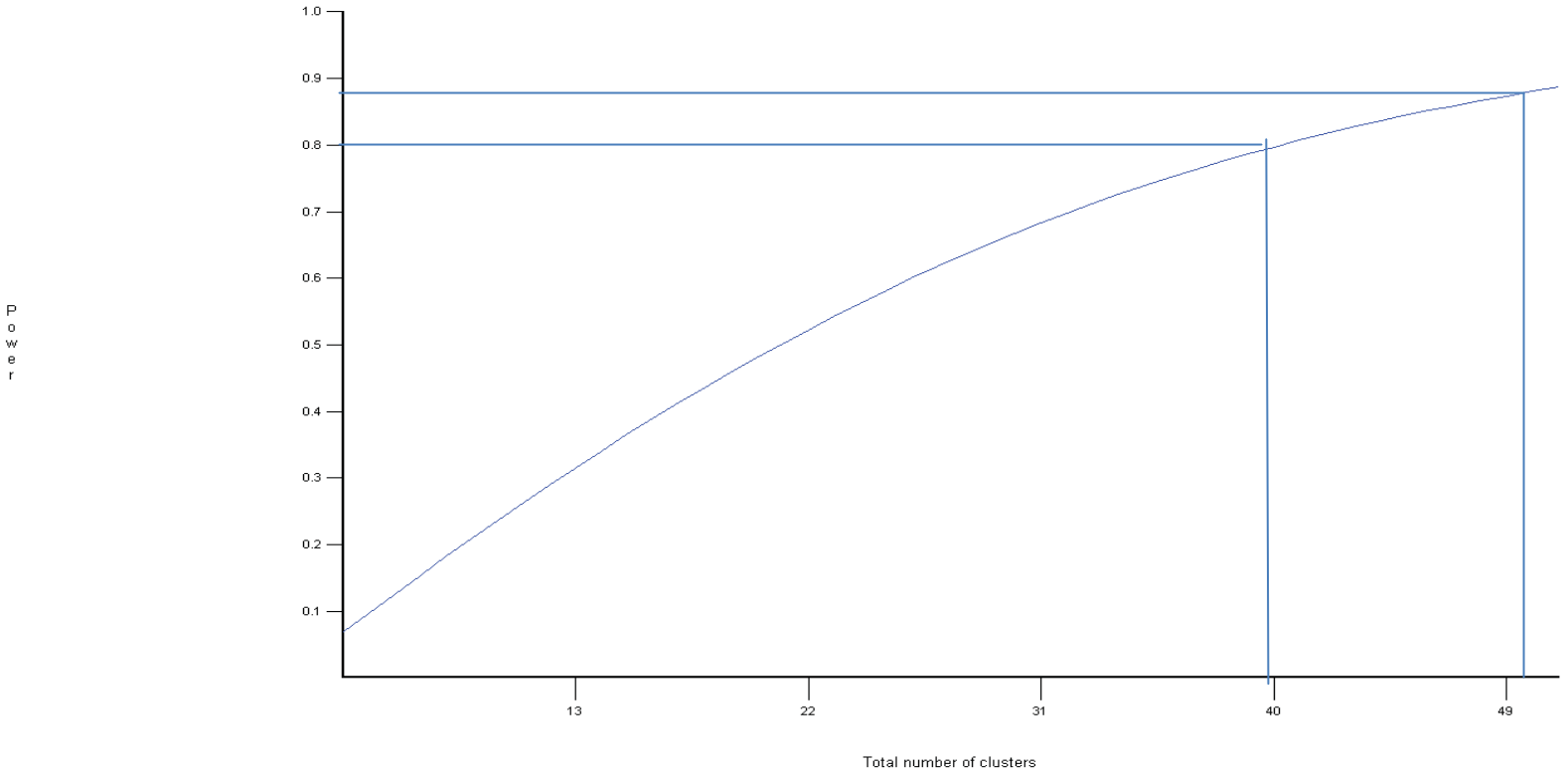

Figure A.2.a

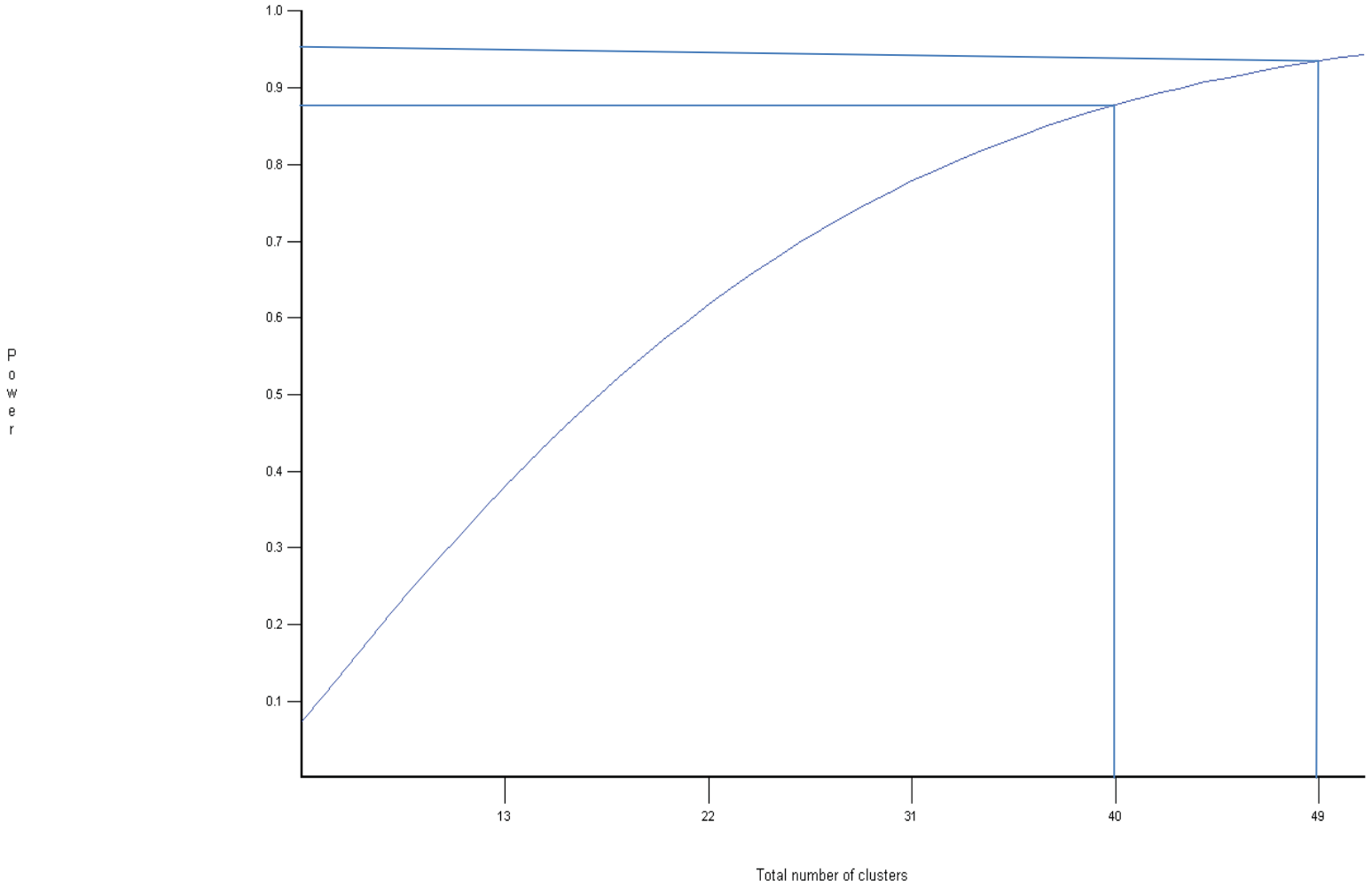
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
Power
13
22
31
40
49
Total number of clusters

Figure A.2.b

Figure A.3: *Power calculation for continuous outcome variable affective attitude.*

From empirical analyses: ICC = 0.006

Assuming effect size is 0.2, proportion of the variance explained by group level covariates: 0.4. Number of subjects in each platoon is 28 (participation rate=50%, Figure A.3.a) to 39 (participation rate=70%, Figure A.3.b). Significant level: a=0.05.

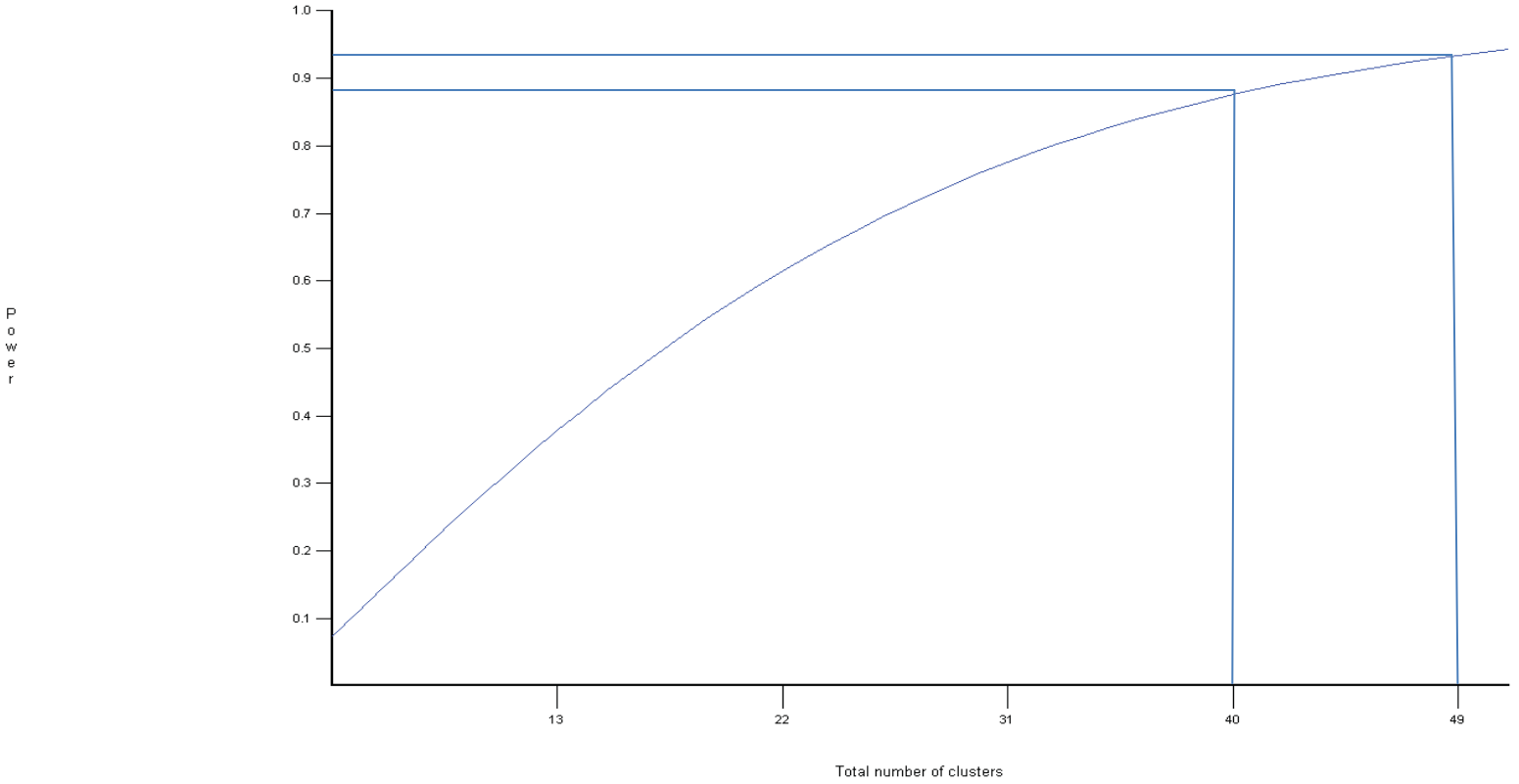

Figure A.3.a

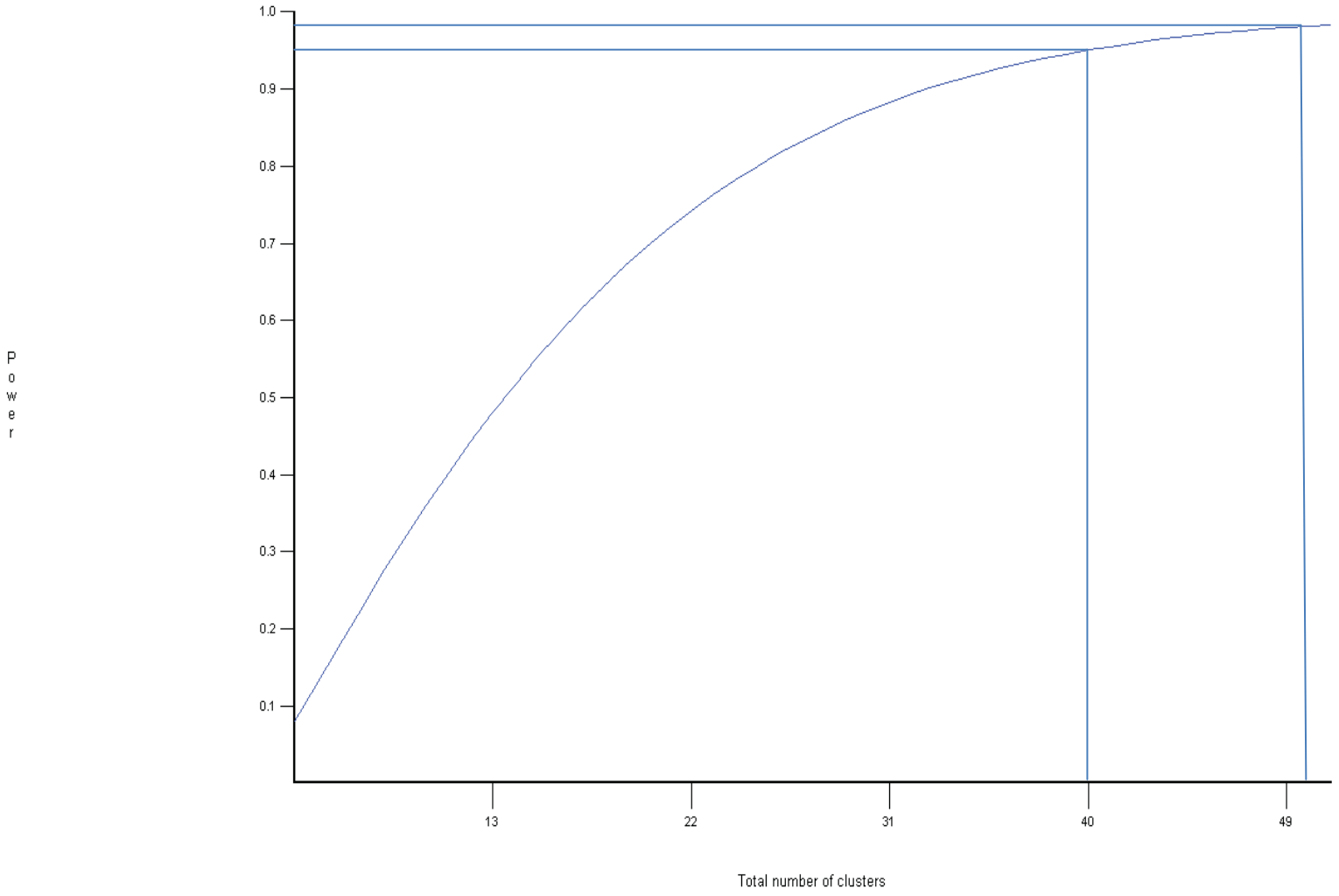
Power
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
13
22
31
40
49
Total number of clusters

Figure A.3.b

Figure A.4: *Power calculation for continuous outcome variable overall intention.*

From empirical analyses: ICC = 0.008

Assuming effect size is 0.2, proportion of the variance explained by group level covariates: 0.4. Number of subjects in each platoon is 28 (participation rate=50%, Figure A.4.a) to 39 (participation rate=70%, Figure A.4.b). Significant level: a=0.05.

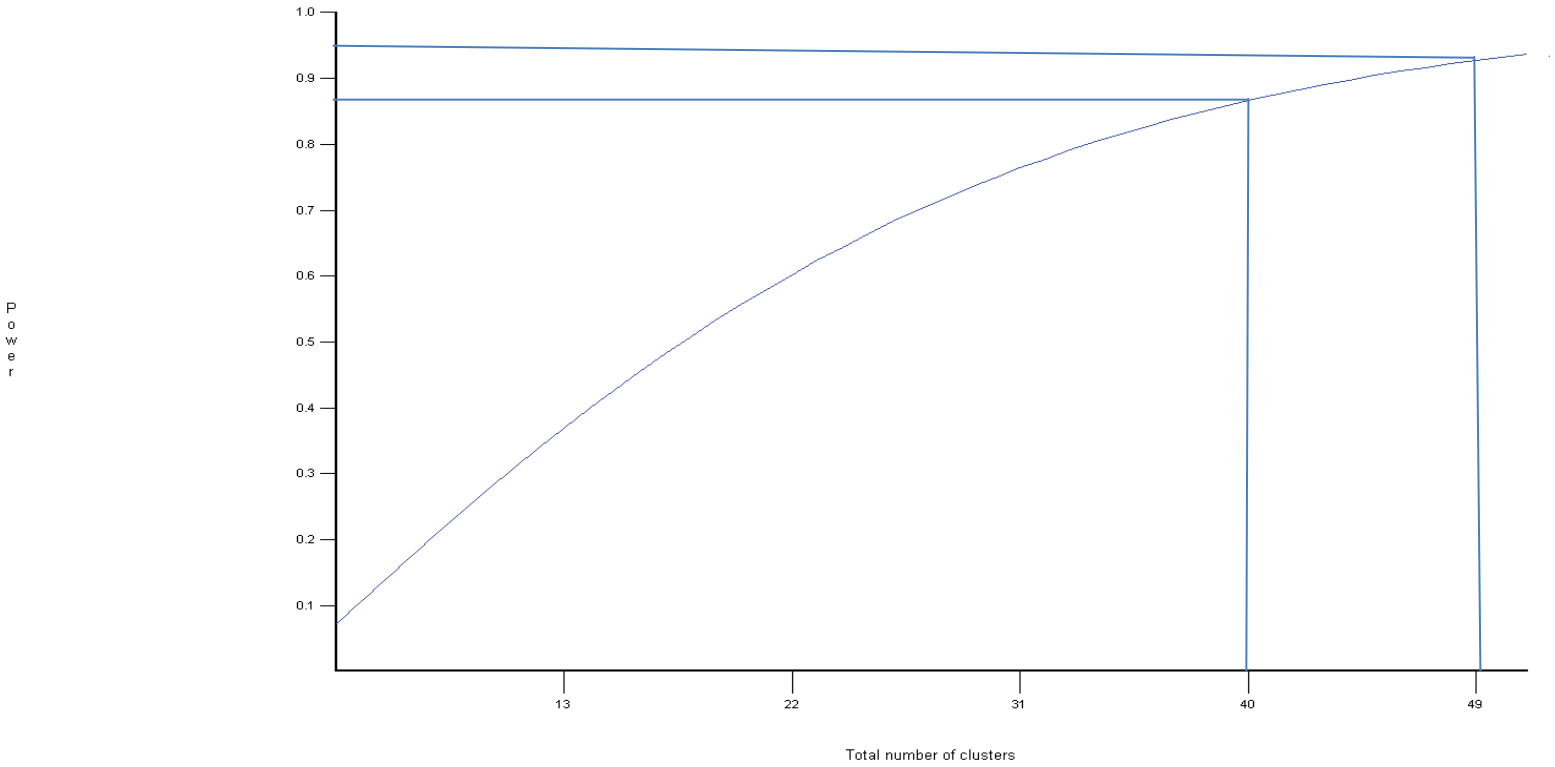

Figure A.4.a

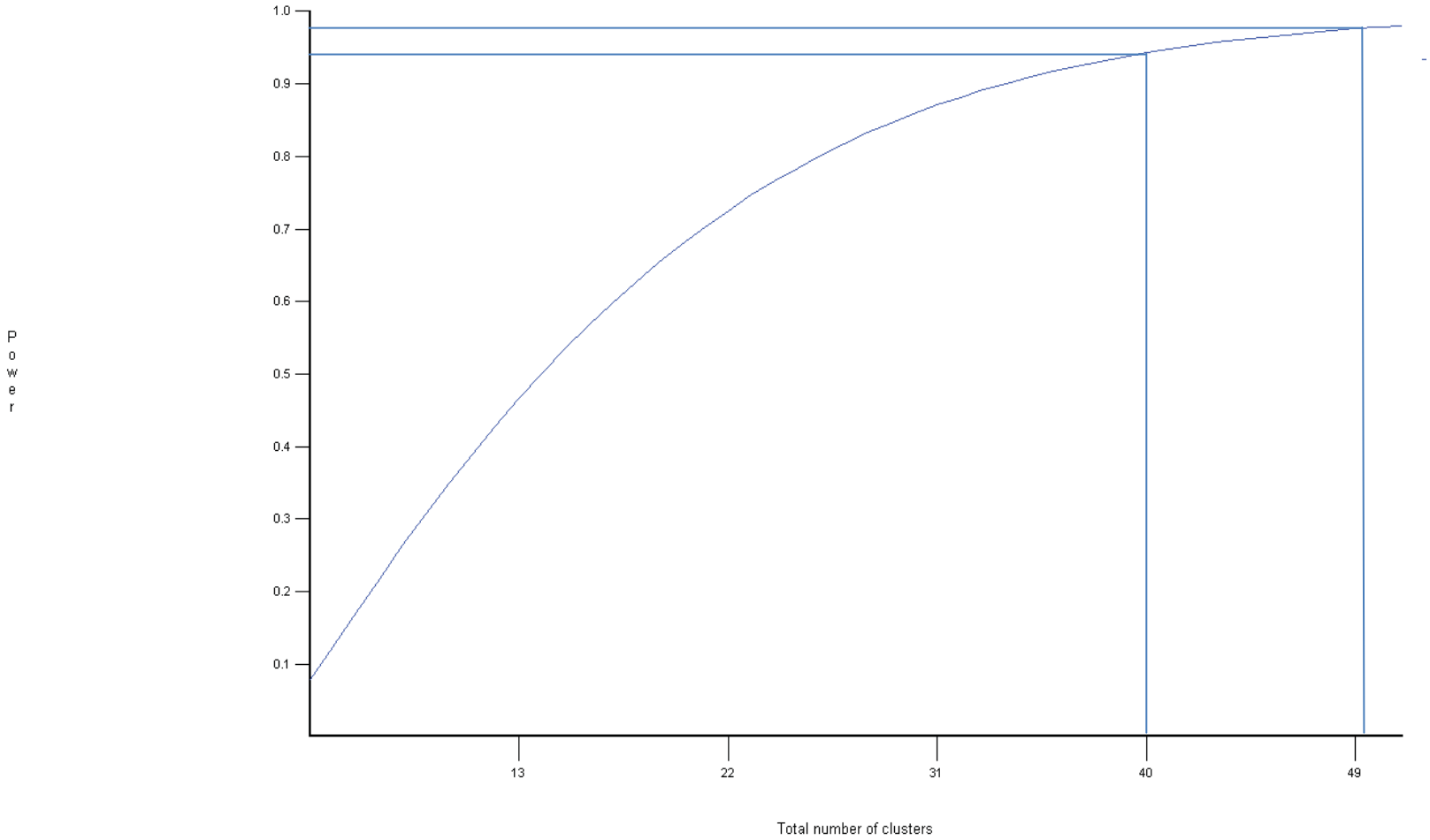
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
Power
13
22
31
40
49
Total number of clusters

Figure A.4.b

Figure A.5: *Power calculation for continuous outcome variable overall perceived norms.*

From empirical analyses: ICC = 0.009

Assuming effect size is 0.2, proportion of the variance explained by group level covariates: 0.4. Number of subjects in each platoon is 28 (participation rate=50%, Figure A.5.a) to 39 (participation rate=70%, Figure A.5.b). Significant level: a=0.05.

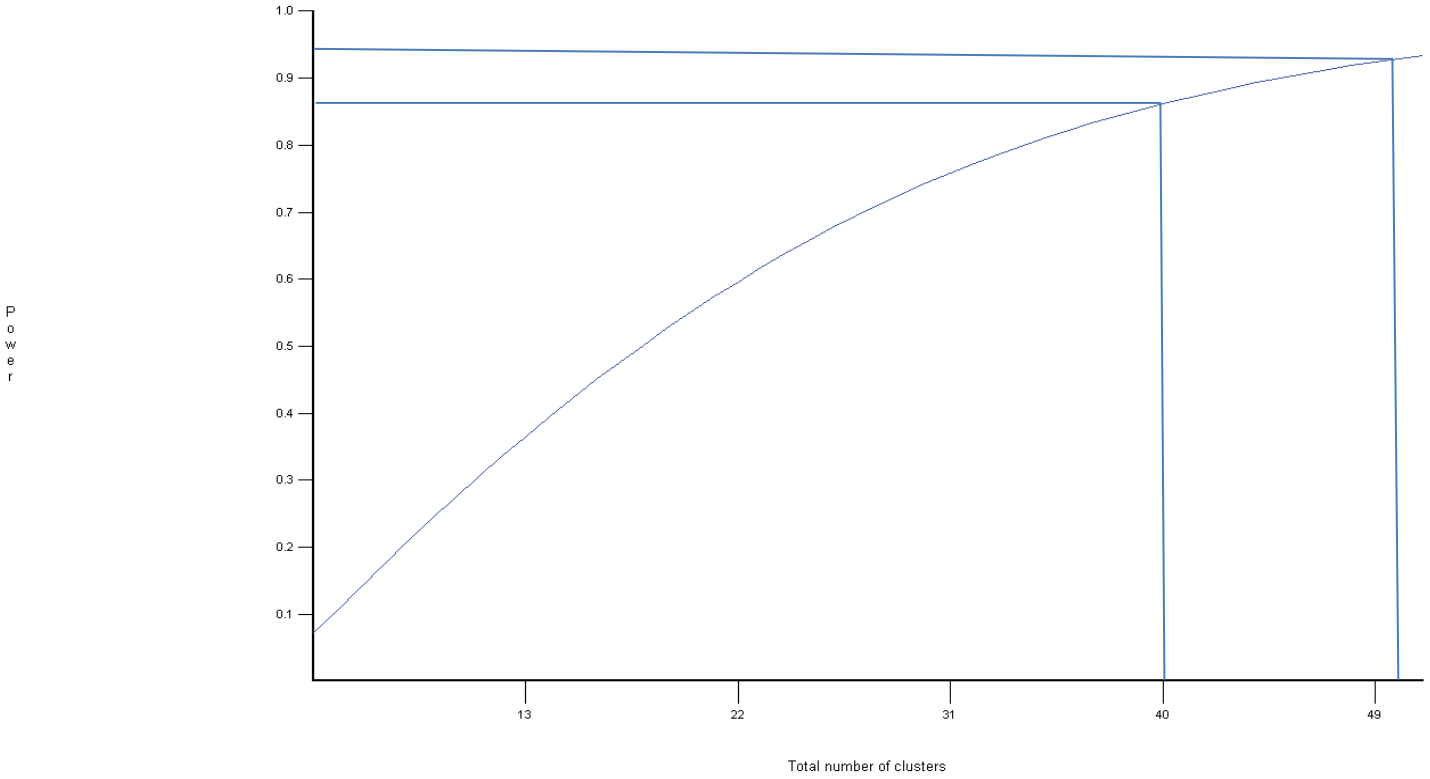

Figure A.5.a

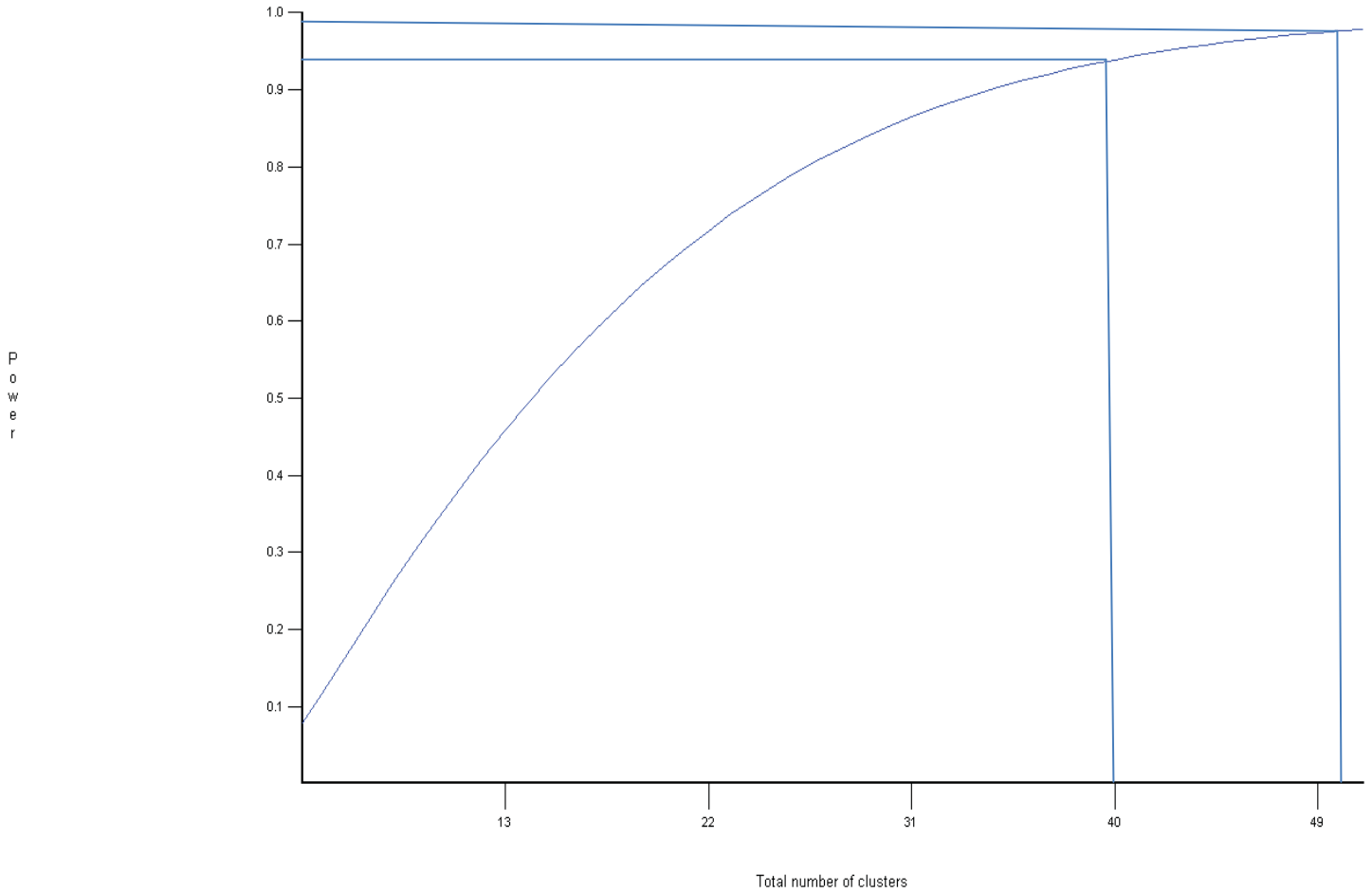
Power
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
13
22
31
40
49
Total number of clusters

Figure A.5.b

Figure A.6: *Power calculation for continuous outcome variable overall perceived control.*

From empirical analyses: ICC = 0.027

Assuming effect size is 0.2, proportion of the variance explained by group level covariates: 0.4. Number of subjects in each platoon is 28 (participation rate=50%, Figure A.6.a) to 39 (participation rate=70%, Figure A.6.b). Significant level: a=0.05.

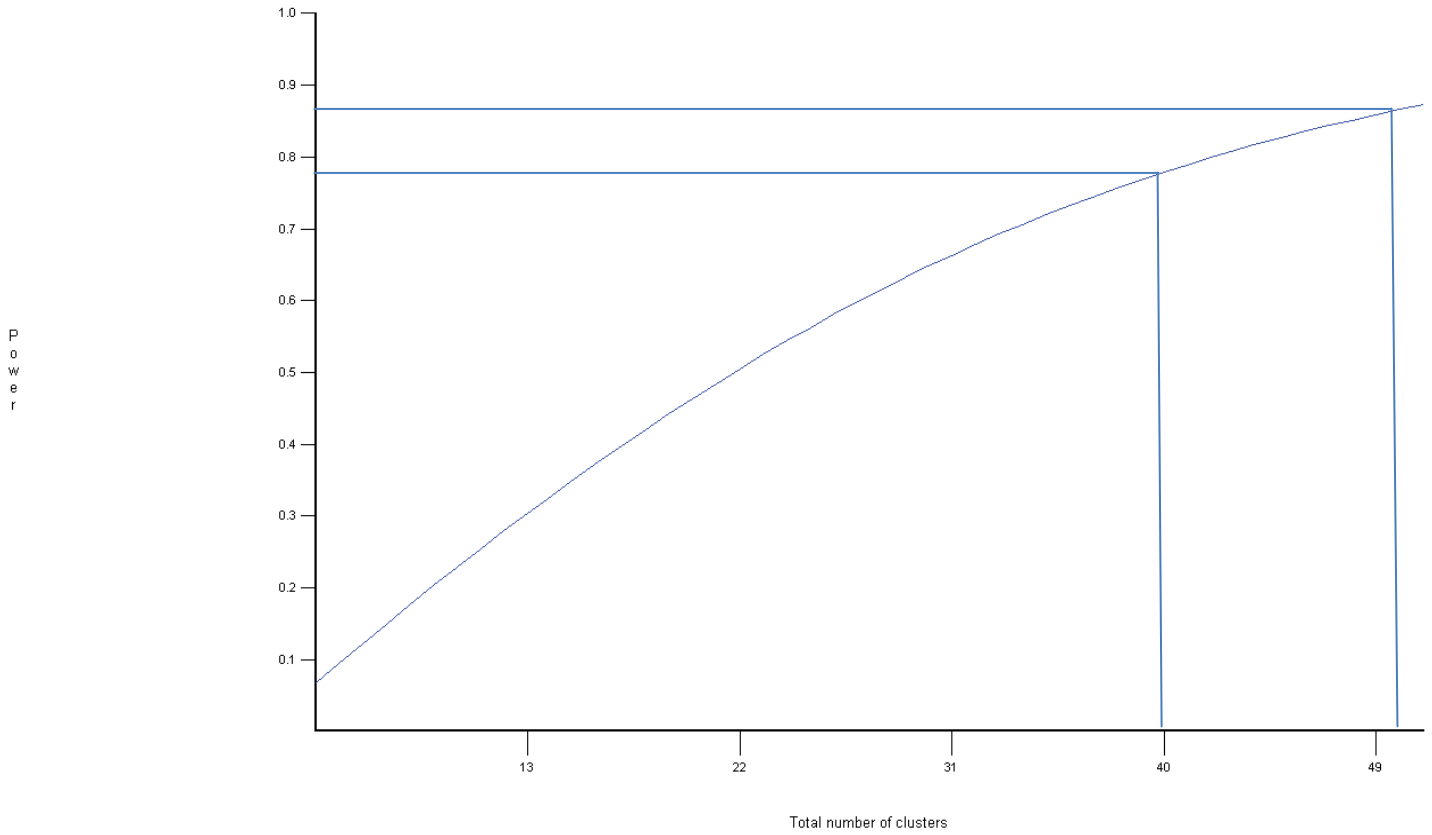

Figure A.6.a

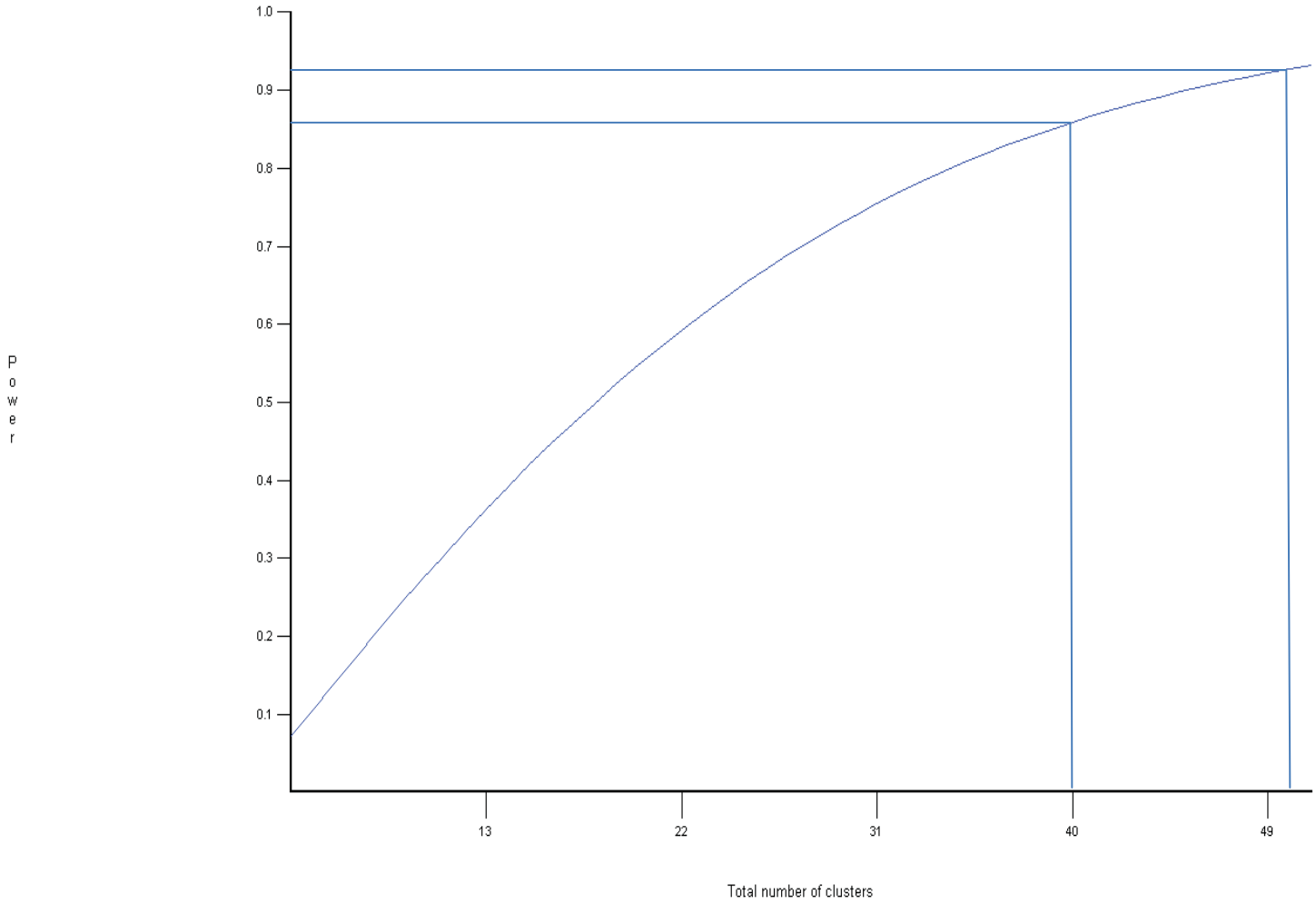
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
Power
13
22
31
40
49
Total number of clusters

Figure A.6.b

Figure A.7: *Power calculation for continuous outcome variable perceived control.*

From empirical analyses: ICC = 0

Assuming effect size is 0.2, proportion of the variance explained by group level covariates: 0.4. Number of subjects in each platoon is 28 (participation rate=50%, Figure A.7.a) to 39 (participation rate=70%, Figure A.7.b). Significant level: a=0.05.

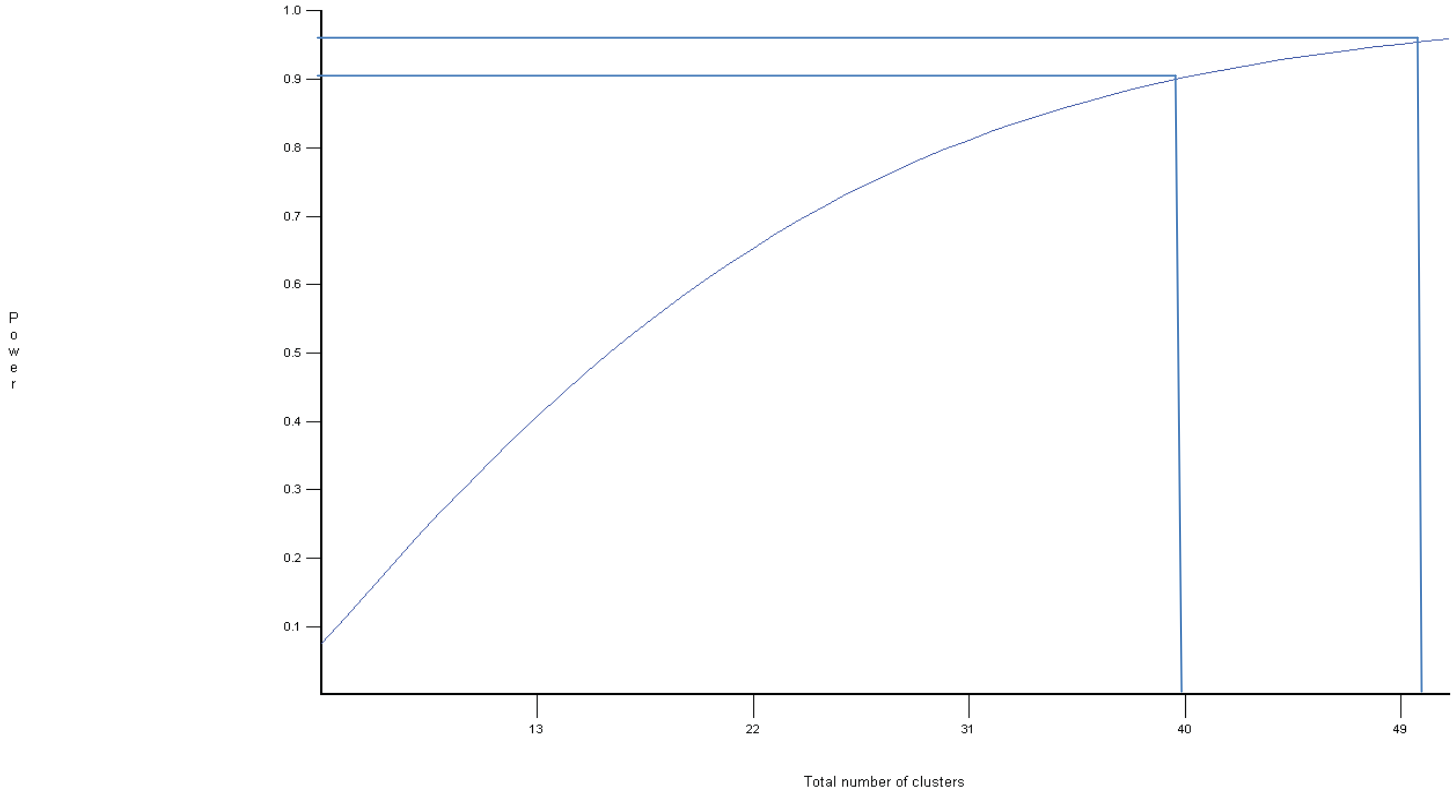

Figure A.7.a

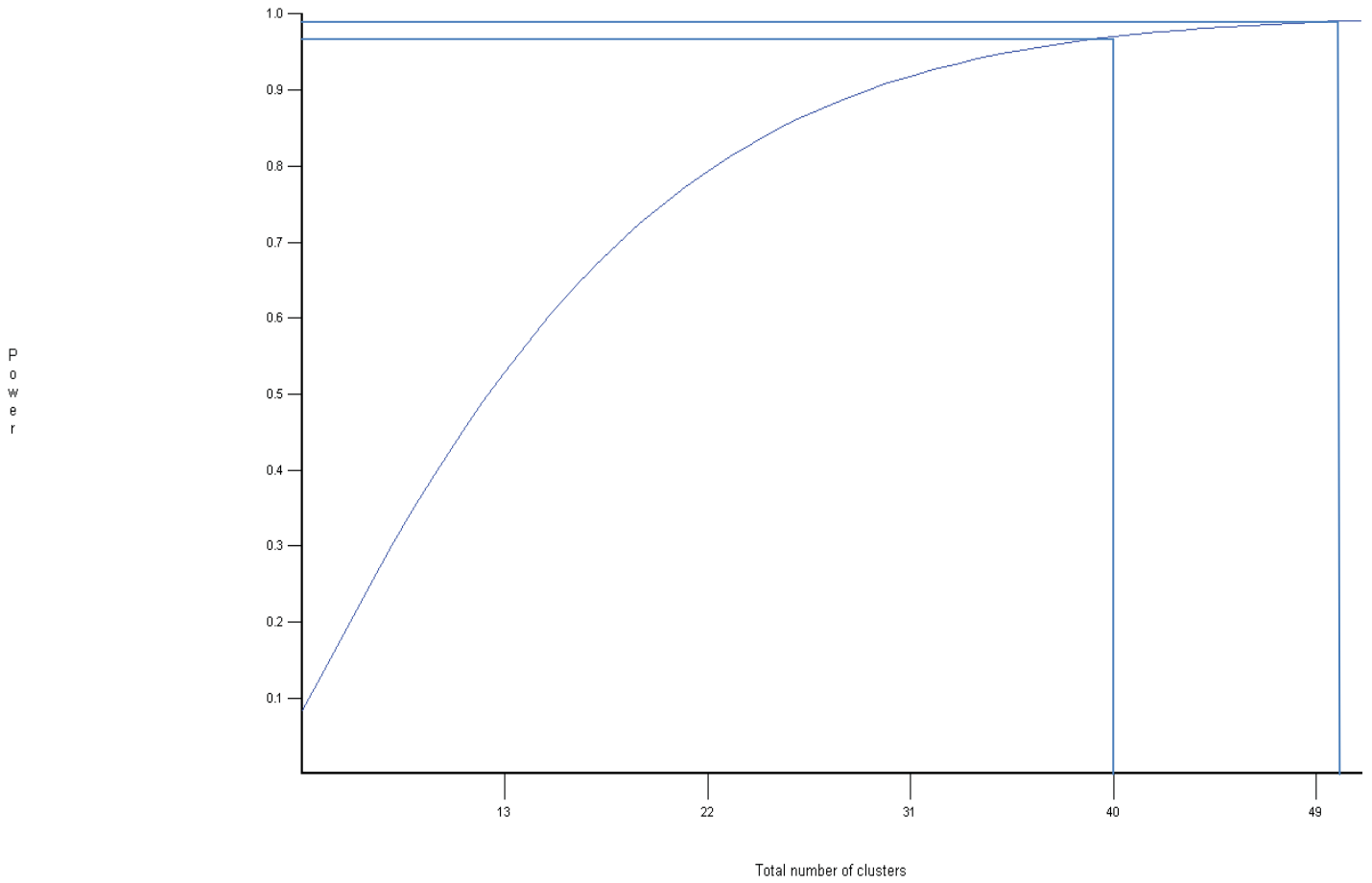
Power
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
13
22
31
40
49
Total number of clusters

Figure A.7.b

Figure A.8: *Power calculation for continuous outcome variable self-efficacy.*

From empirical analyses: ICC = 0.038

Assuming effect size is 0.2, proportion of the variance explained by group level covariates: 0.4. Number of subjects in each platoon is 28 (participation rate=50%, Figure A.8.a) to 39 (participation rate=70%, Figure A.8.b). Significant level: a=0.05.

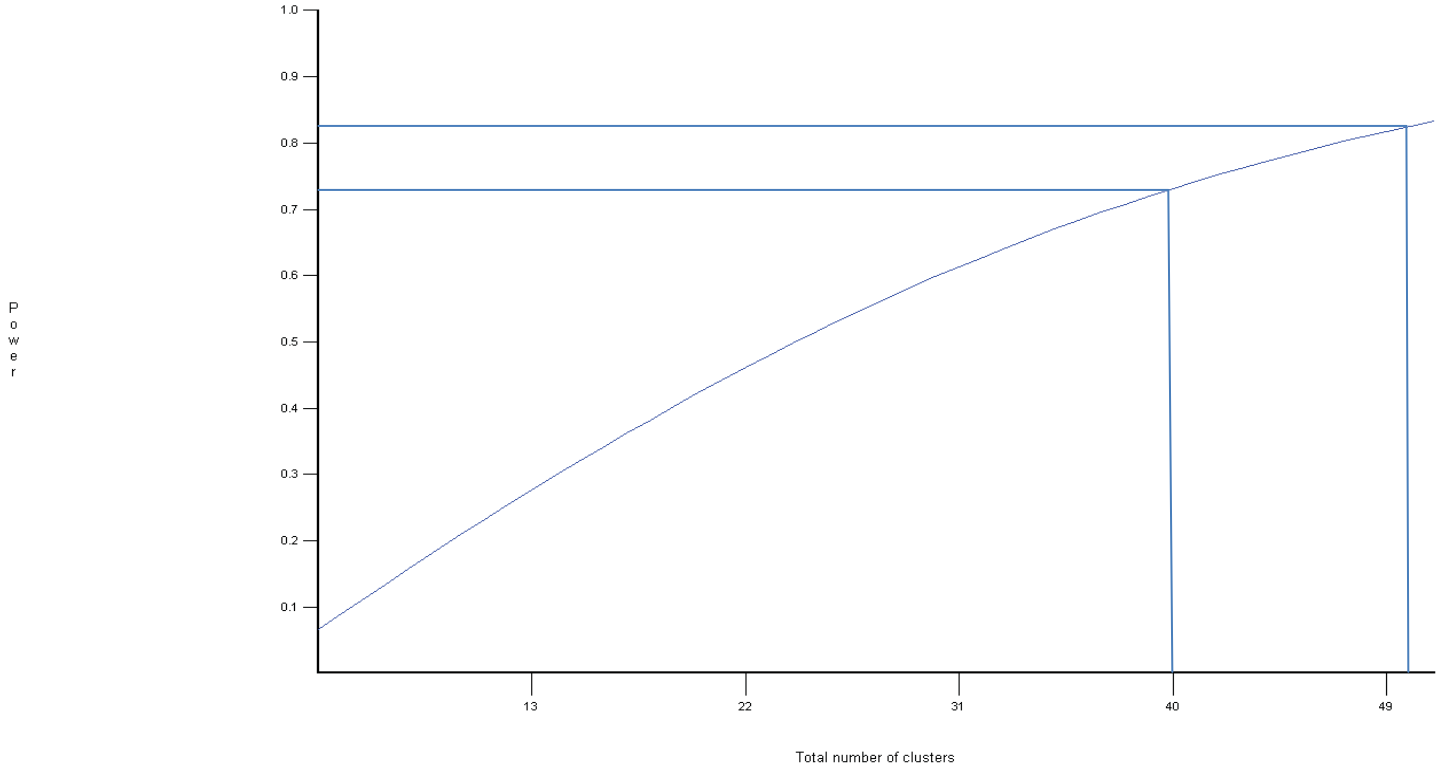

Figure A.8.a

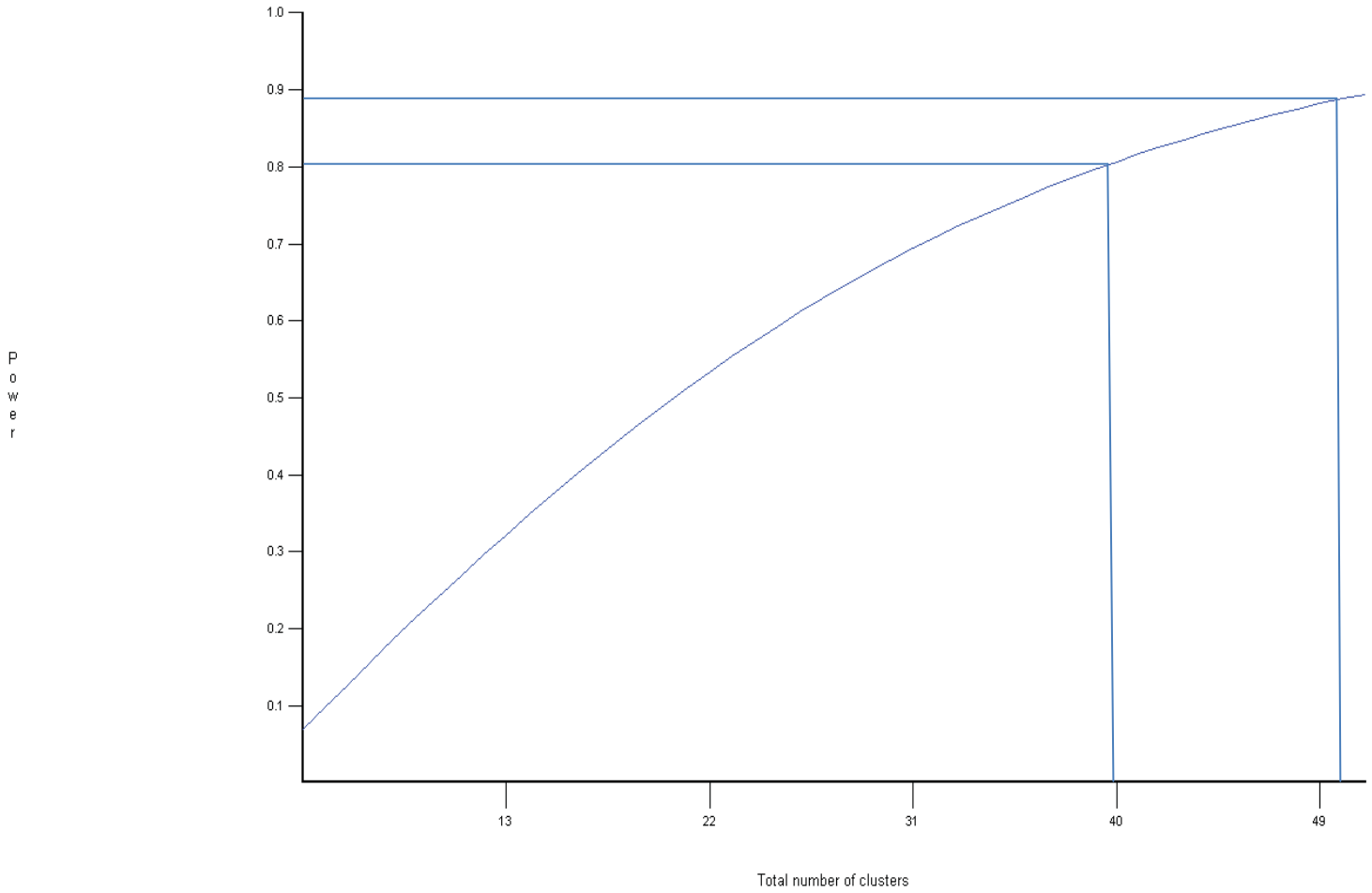
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
Power
13
22
31
40
49
Total number of clusters

Figure A.8.b

Figure A.9: *Power calculation for continuous outcome variable PHQ9.*

From empirical analyses: ICC = 0.007

Assuming effect size is 0.2, proportion of the variance explained by group level covariates: 0.4. Number of subjects in each platoon is 28 (participation rate=50%, Figure A.9.a) to 39 (participation rate=70%, Figure A.9.b). Significant level: a=0.05.

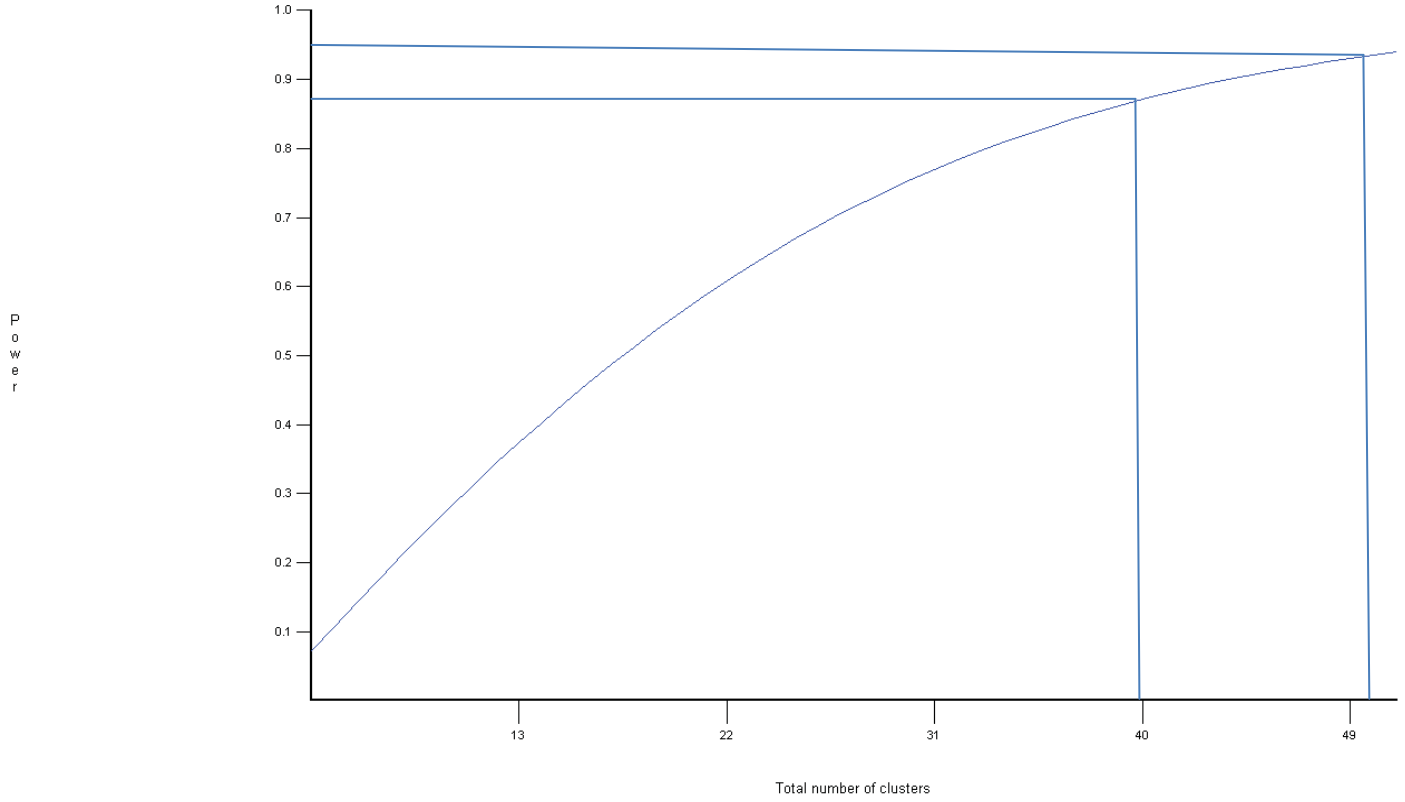

Figure A.9.a

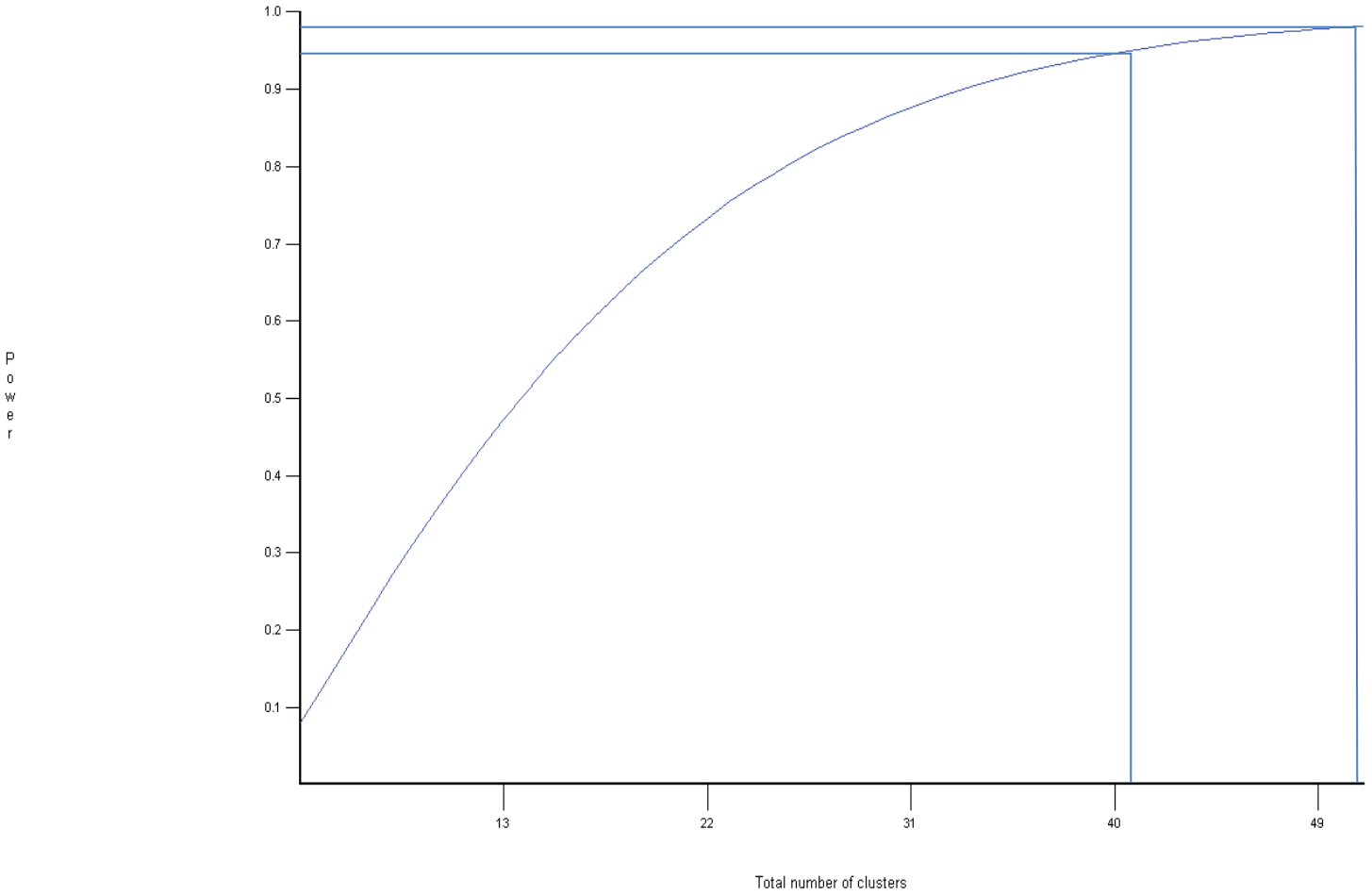
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
Power
13
22
31
40
49
Total number of clusters

Figure A.9.b

Figure A.10: *Power calculation for continuous outcome variable PHQ15.*

From empirical analyses: ICC = 0.025

Assuming effect size is 0.2, proportion of the variance explained by group level covariates: 0.4. Number of subjects in each platoon is 28 (participation rate=50%, Figure A.10.a) to 39 (participation rate=70%, Figure A.10.b). Significant level: a=0.05.

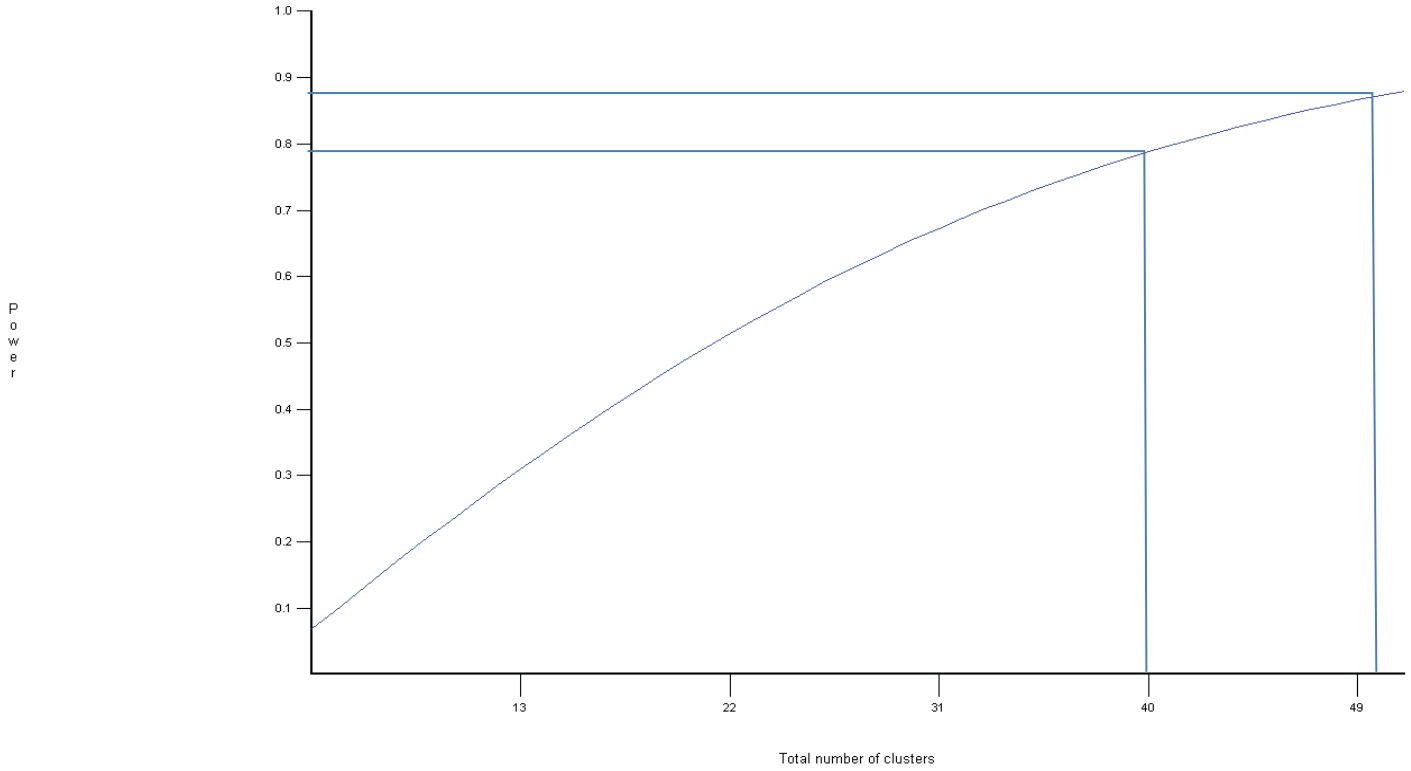

Figure A.10.a

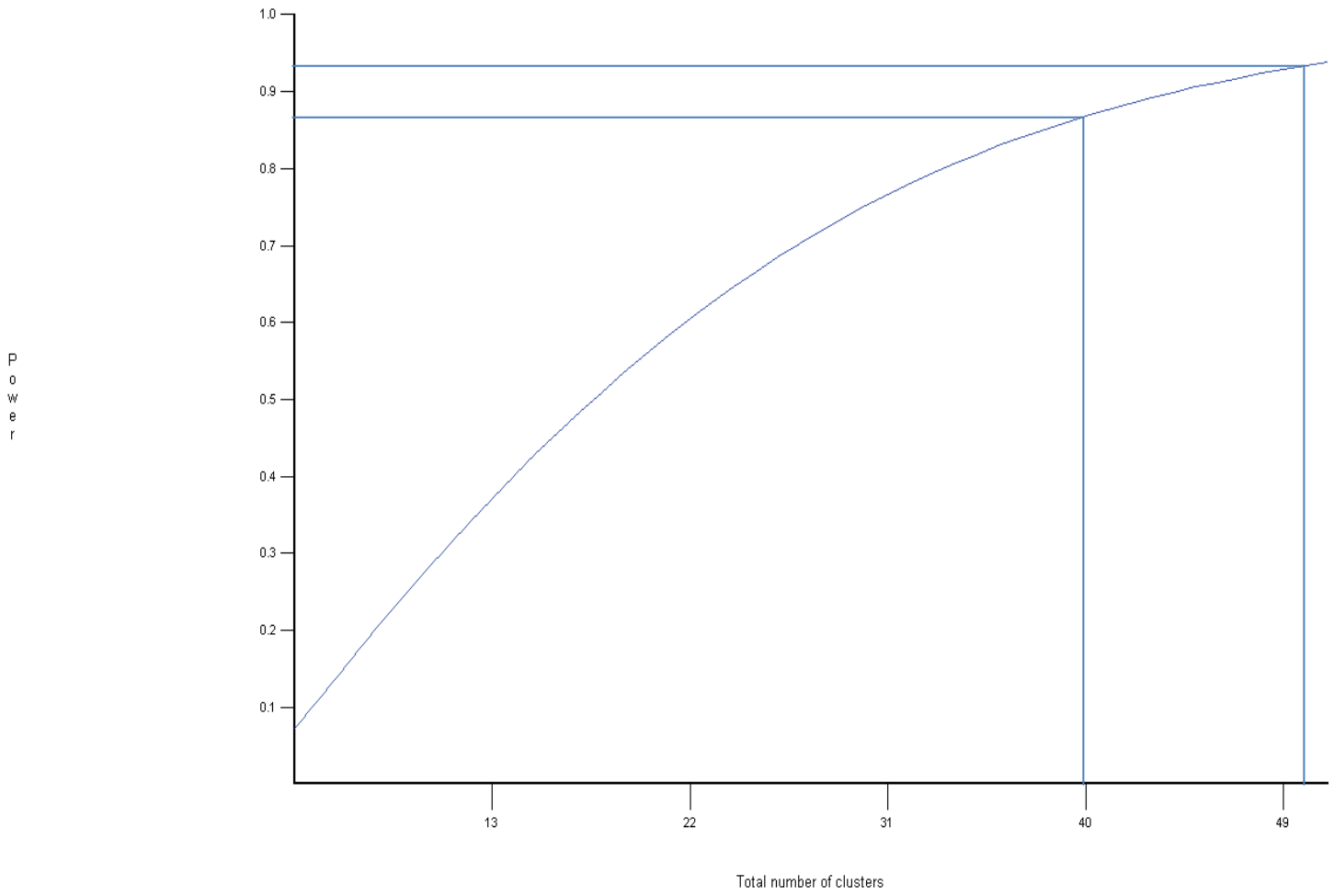
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
Power
13
22
31
40
49
Total number of clusters

Figure A.10.b

Figure A.11: *Power calculation for binary outcome variable BMQ graduation.*

From empirical analyses: BMQ graduation rate = 0.8

Assuming the success rate among control and intervention groups are 0.80 and 0.90, respectively. 95% confidence interval for the success rate in control group is 0.75 – 0.85. Number of subjects in each platoon is 21 (participation rate=50%, dropout rate=15%, Figure A.11.a) to 29 (participation rate=70%, dropout rate=15%, Figure A.11.b).

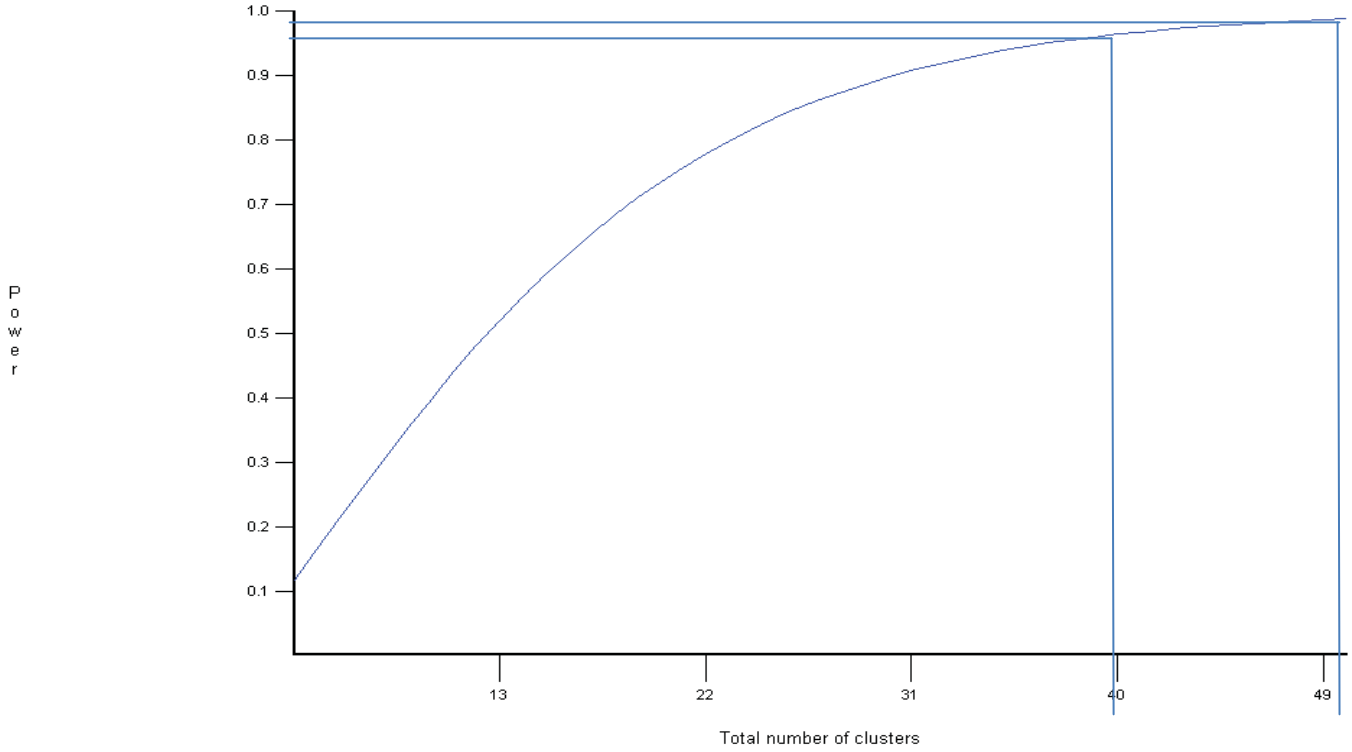

Figure A.11.a

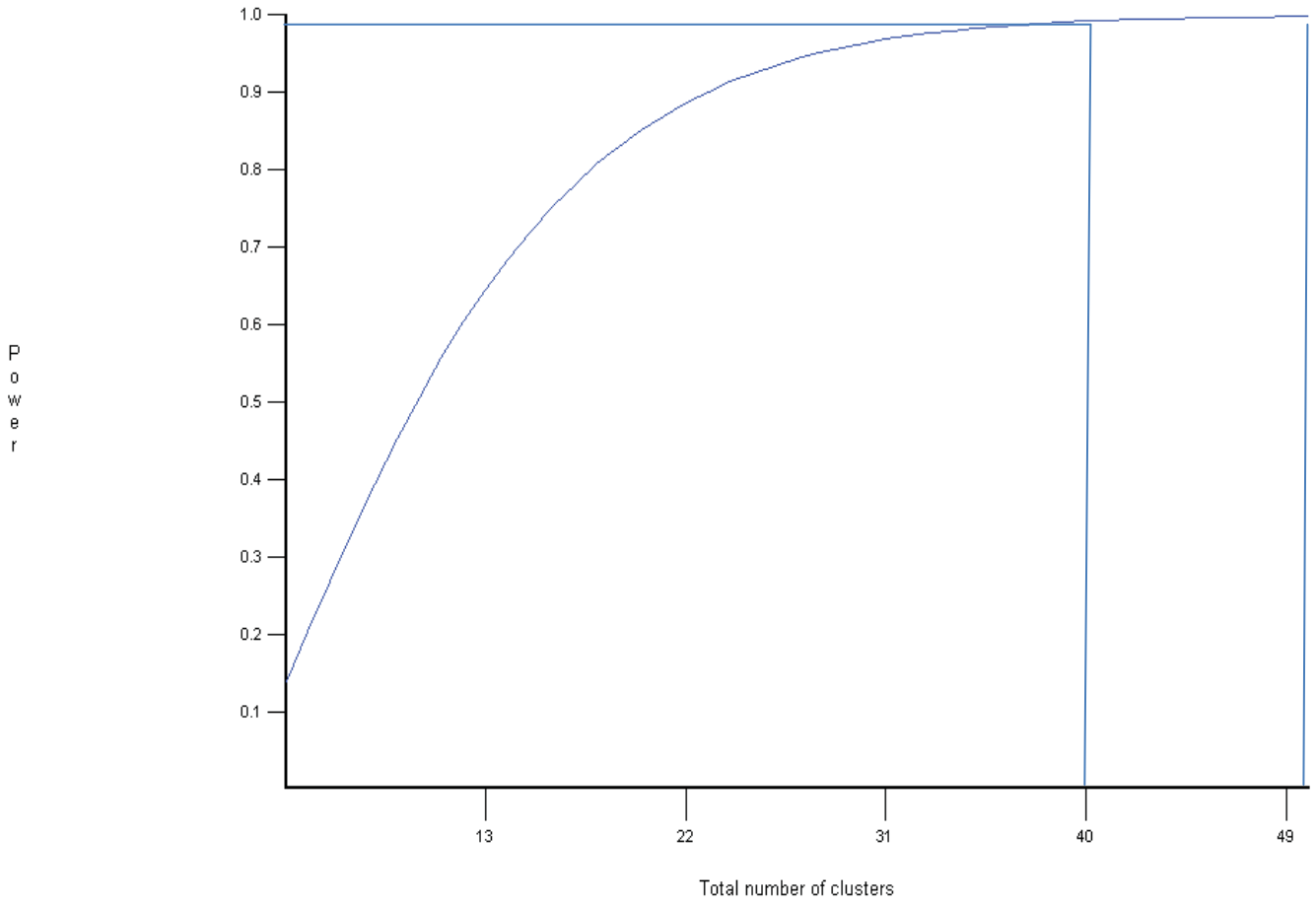
Power
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
13
22
31
40
49
Total number of clusters

Figure A.11.b

Figure A.12: *Minimum detectable effect size calculation for continuous outcome variable overall attitude.*

From empirical analyses: ICC = 0.015

Assuming power=0.8, proportion of the variance explained by group level covariates: 0.4. Number of subjects in each platoon is 28 (participation rate=50%, Figure A.12.a) to 39 (participation rate=70%, Figure A.12.b). Significant level: a=0.05.

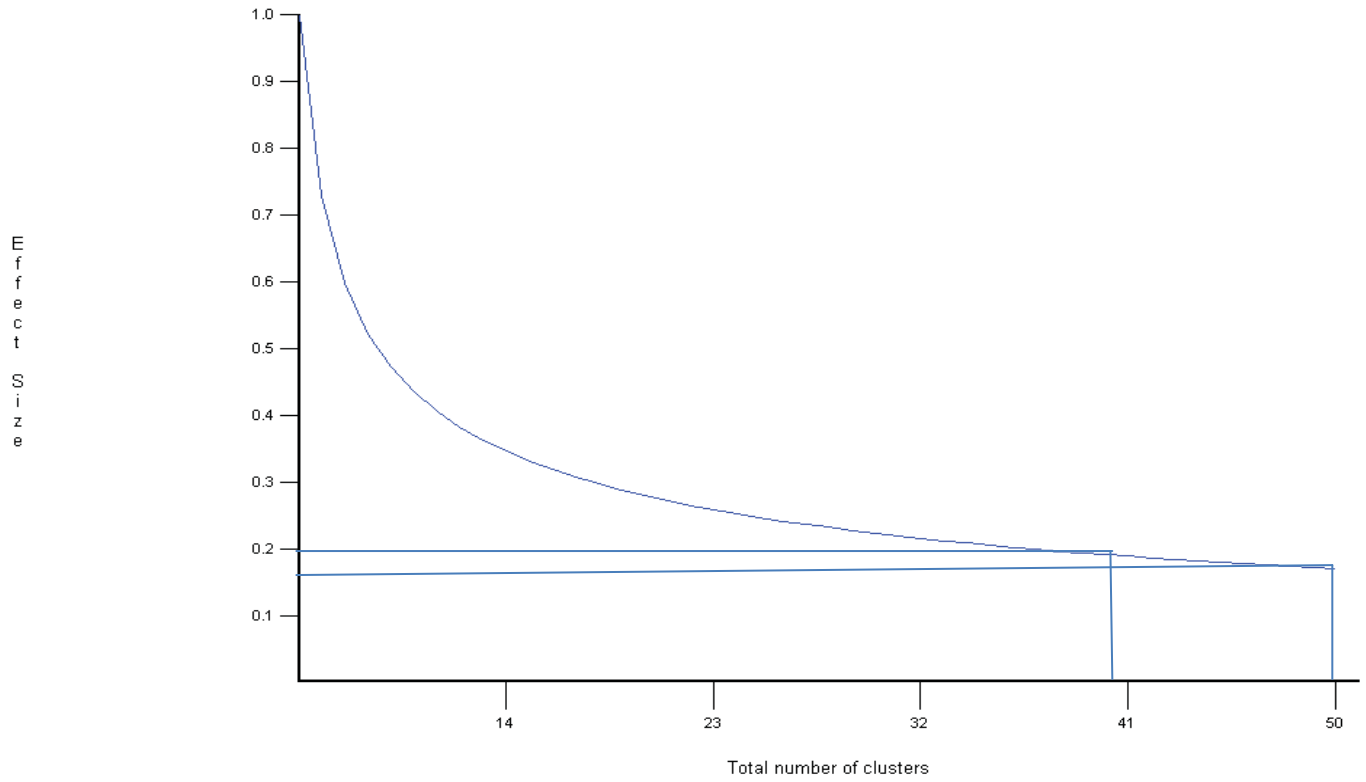

Figure A.12.a

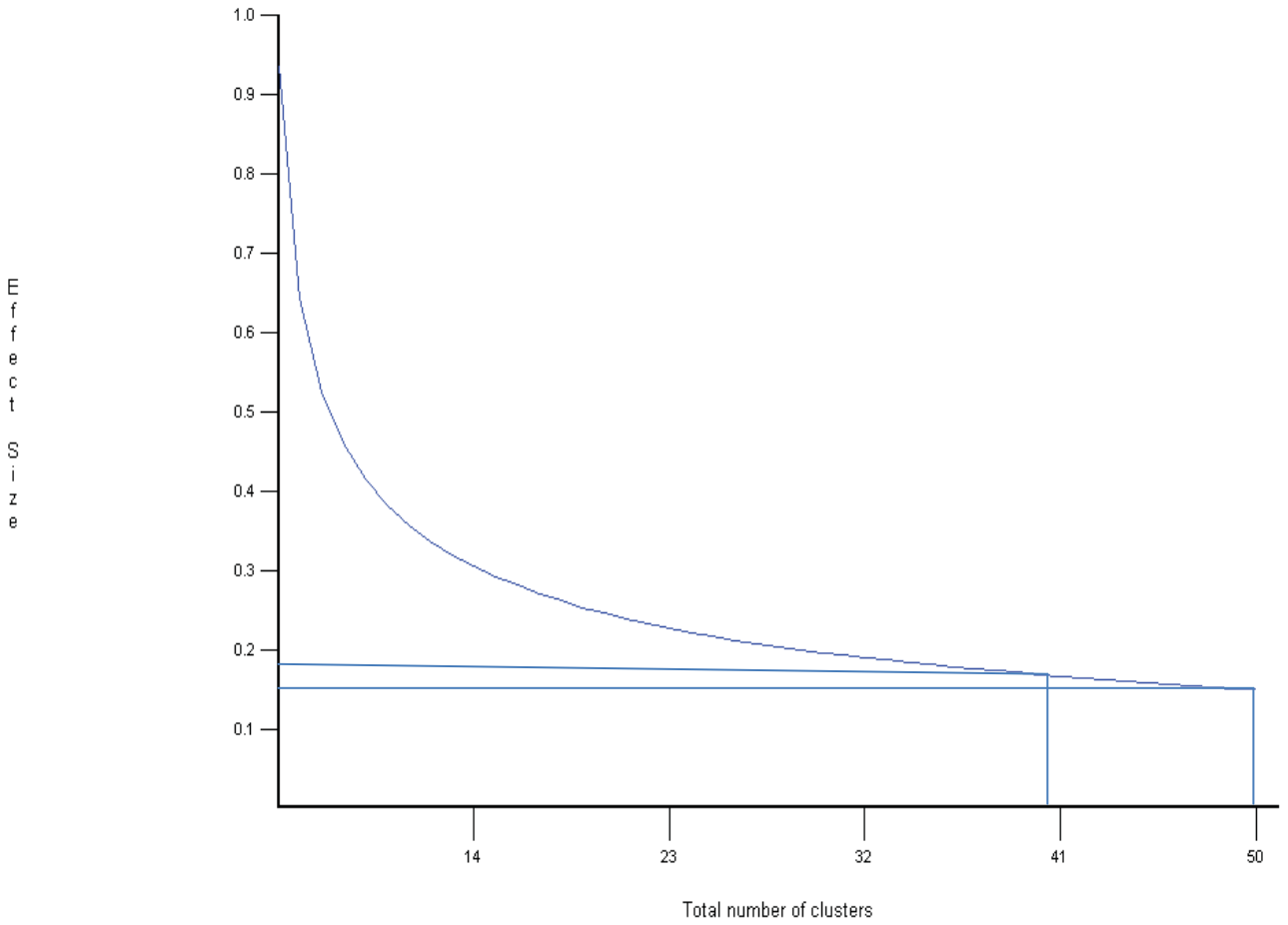
Effect Size
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
14
23
32
41
50
Total number of clusters

Figure A.12.b

Figure A.13: *Minimum detectable effect size calculation for continuous outcome variable instrumental attitude.*

From empirical analyses: ICC = 0.023

Assuming power=0.8, proportion of the variance explained by group level covariates: 0.4. Number of subjects in each platoon is 28 (participation rate=50%, Figure A.13.a) to 39 (participation rate=70%, Figure A.13.b). Significant level: a=0.05.

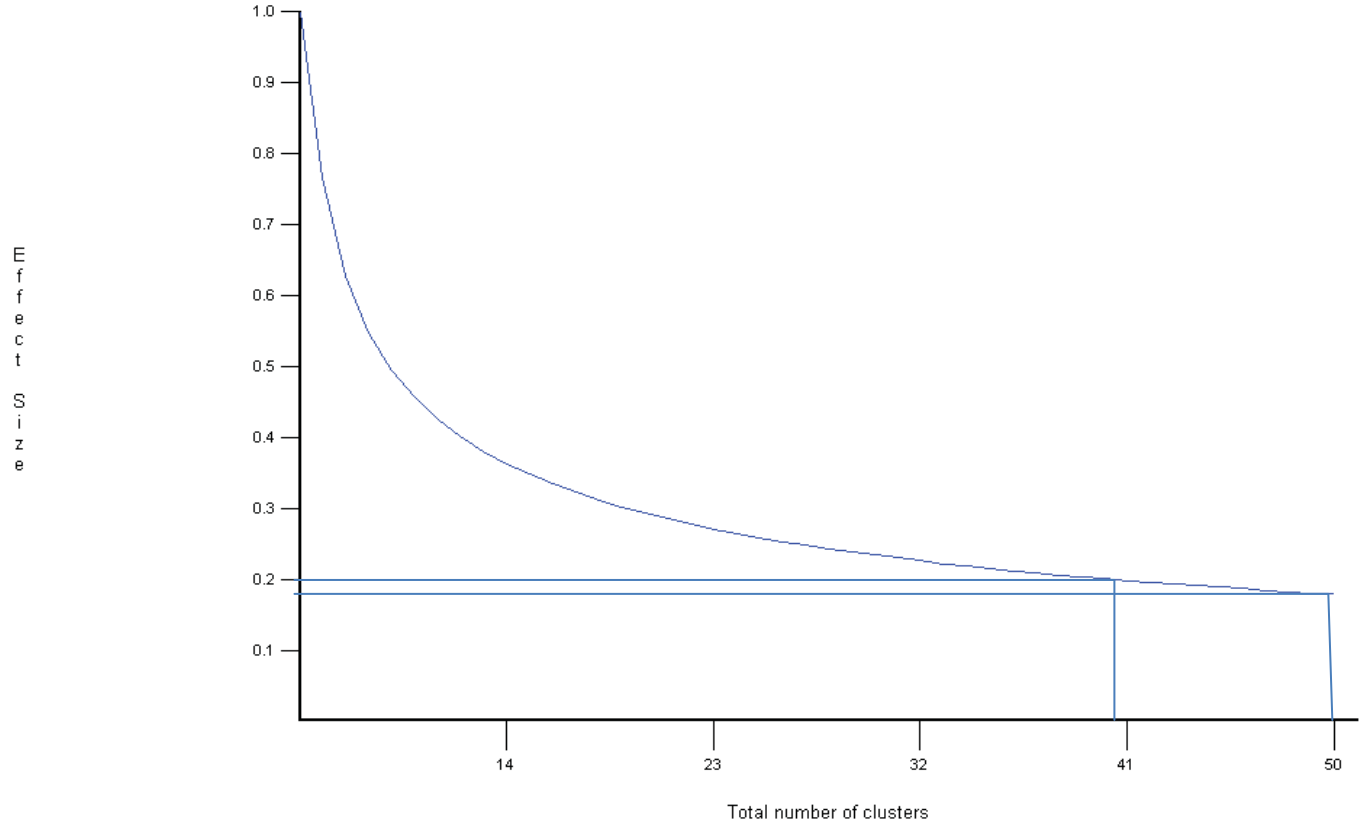

Figure A.13.a

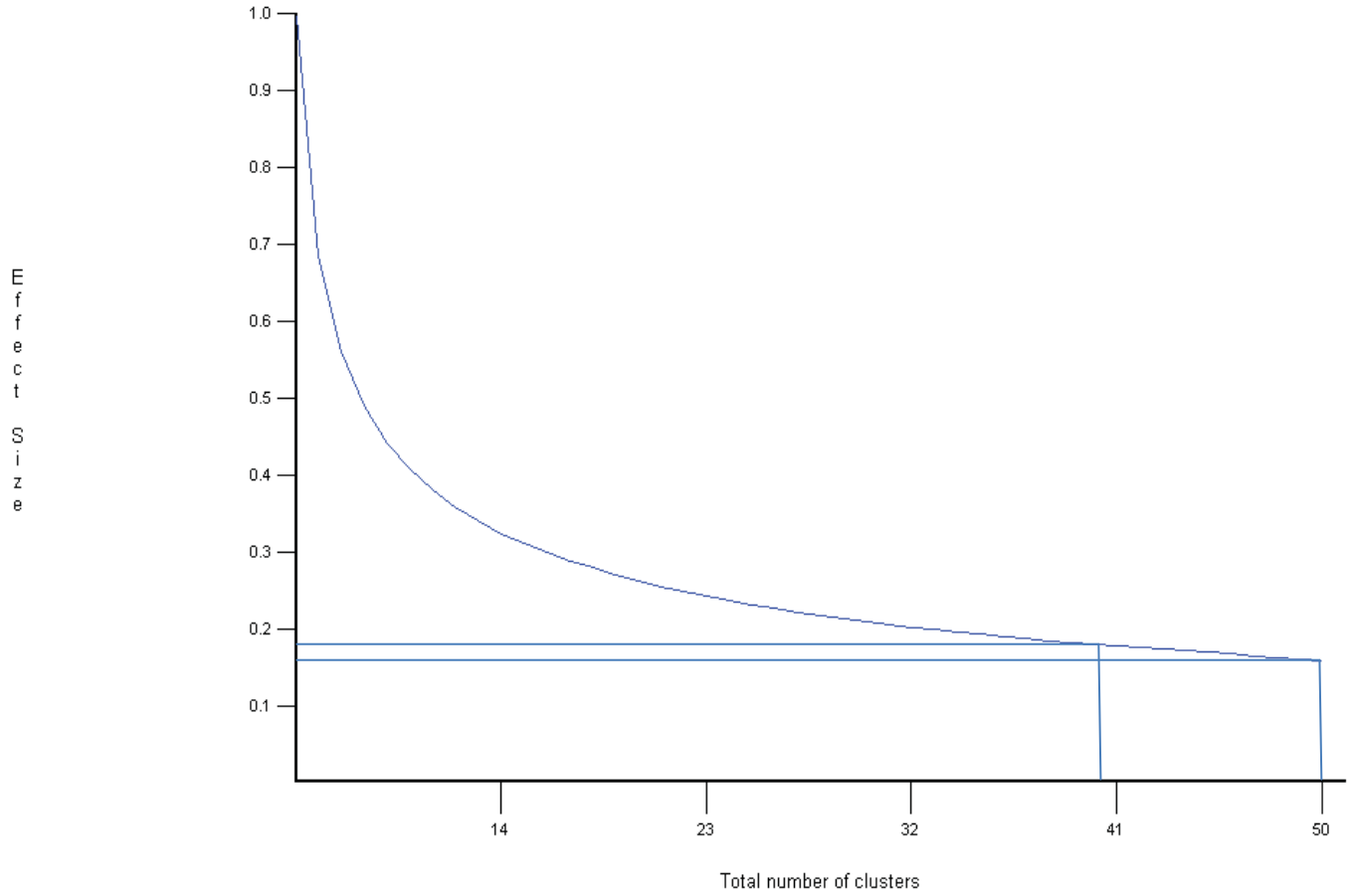
Effect Size
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
14
23
32
41
50
Total number of clusters

Figure A.13.b

Figure A.14: *Minimum detectable effect size calculation for continuous outcome variable affective attitude.*

From empirical analyses: ICC = 0.006

Assuming power=0.8, proportion of the variance explained by group level covariates: 0.4. Number of subjects in each platoon is 28 (participation rate=50%, Figure A.14.a) to 39 (participation rate=70%, Figure A.14.b). Significant level: a=0.05.

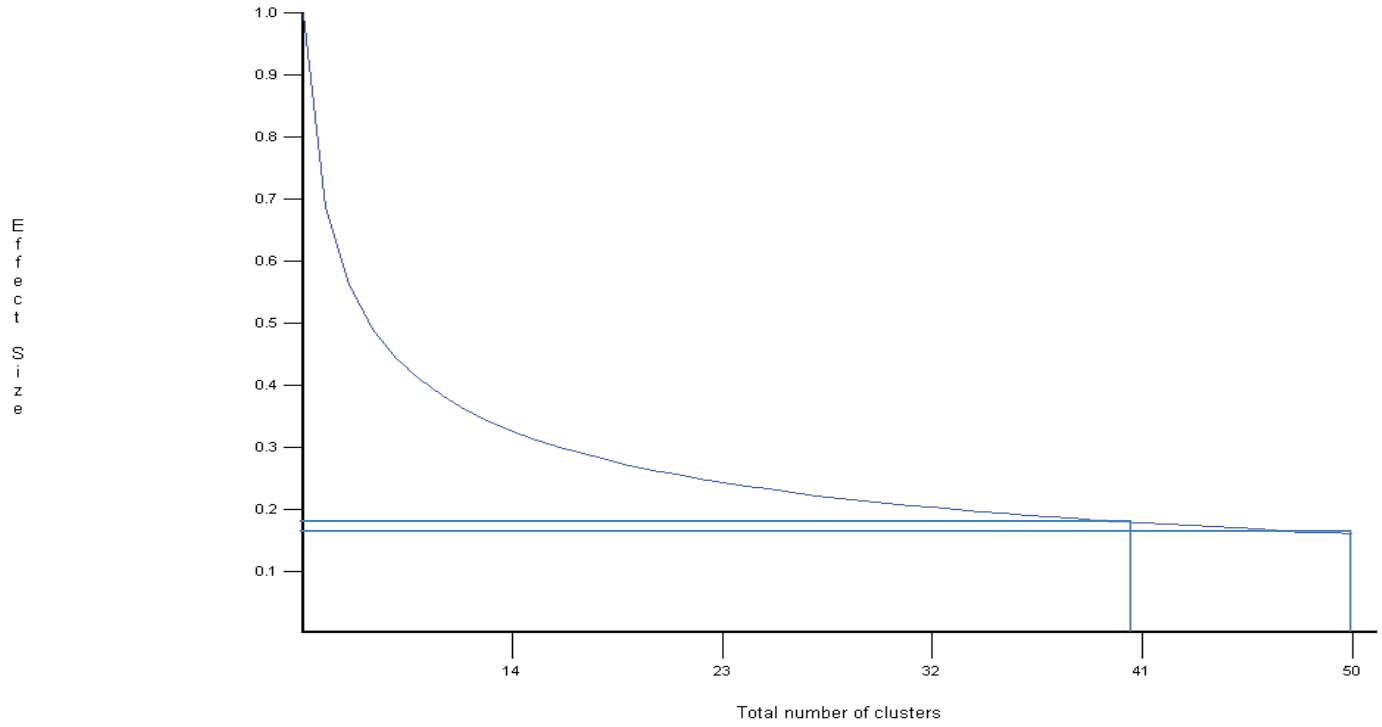

Figure A.14.a

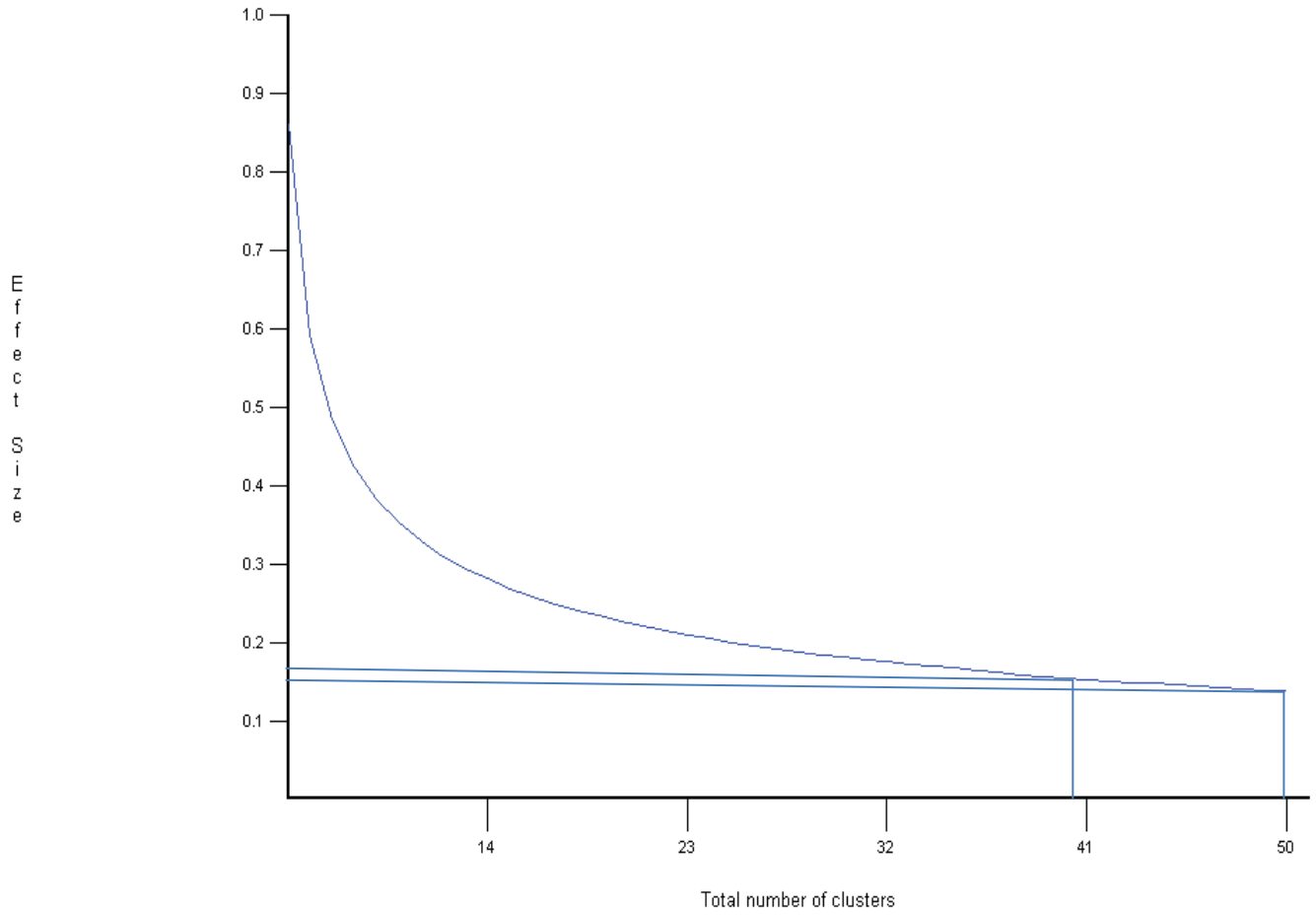
Effect Size
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
14
23
32
41
50
Total number of clusters

Figure A.14.b

Figure A.15: *Minimum detectable effect size calculation for continuous outcome variable overall intention.*

From empirical analyses: ICC = 0.008

Assuming power=0.8, proportion of the variance explained by group level covariates: 0.4. Number of subjects in each platoon is 28 (participation rate=50%, Figure A.15.a) to 39 (participation rate=70%, Figure A.15.b). Significant level: a=0.05.

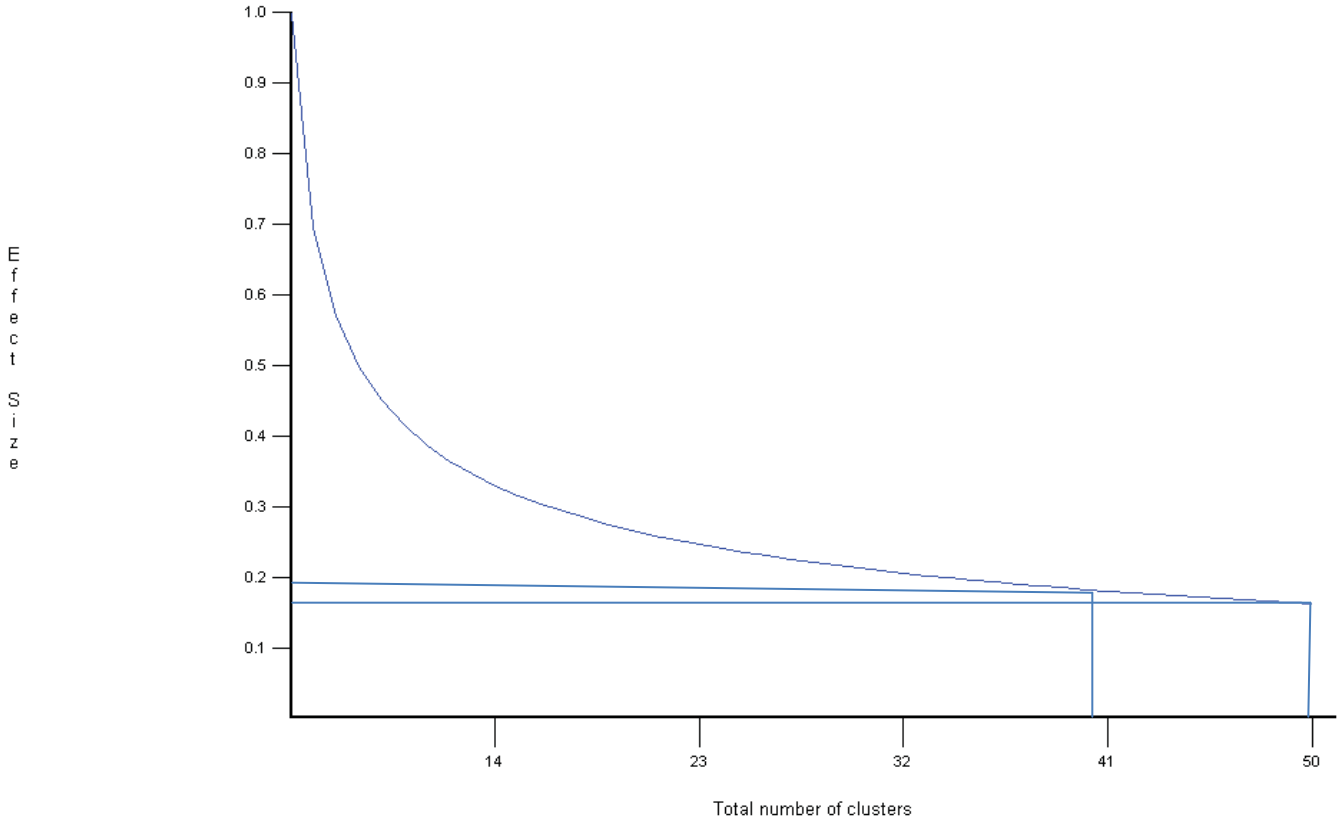

Figure A.15.a

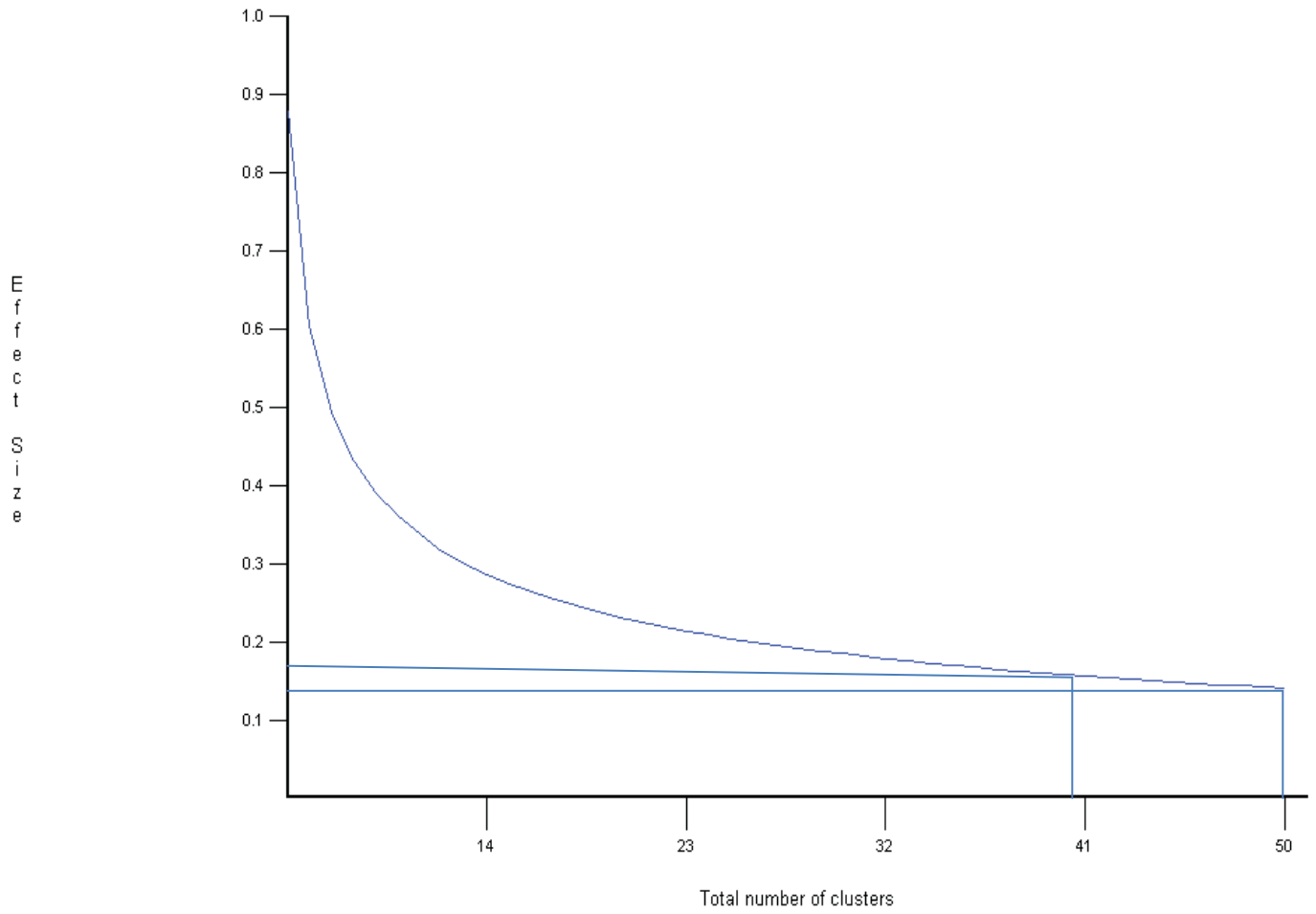
Effect Size
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
14
23
32
41
50
Total number of clusters

Figure A.15.b

Figure A.16: *Minimum detectable effect size calculation for continuous outcome variable overall perceived norms.*

From empirical analyses: ICC = 0.009

Assuming power=0.8, proportion of the variance explained by group level covariates: 0.4. Number of subjects in each platoon is 28 (participation rate=50%, Figure A.16.a) to 39 (participation rate=70%, Figure A.16.b). Significant level: a=0.05.

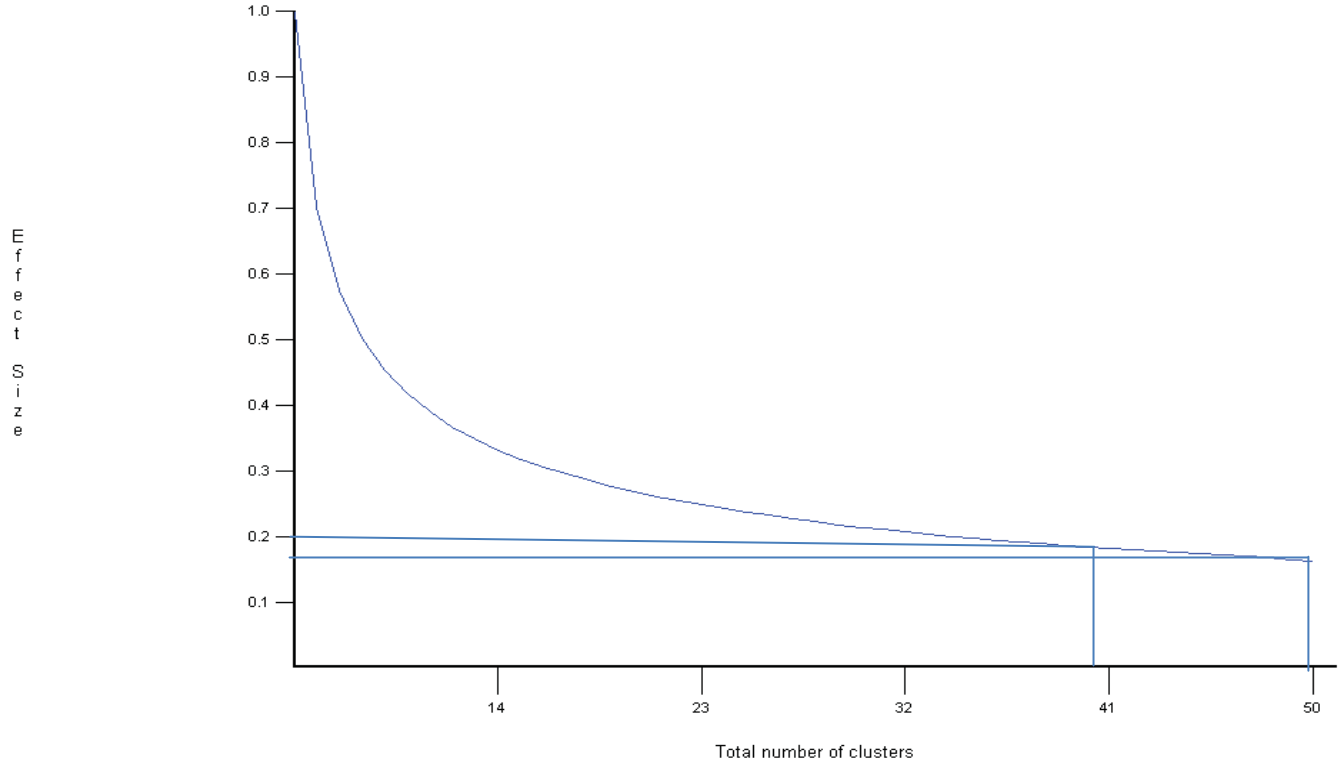

Figure A.16.a

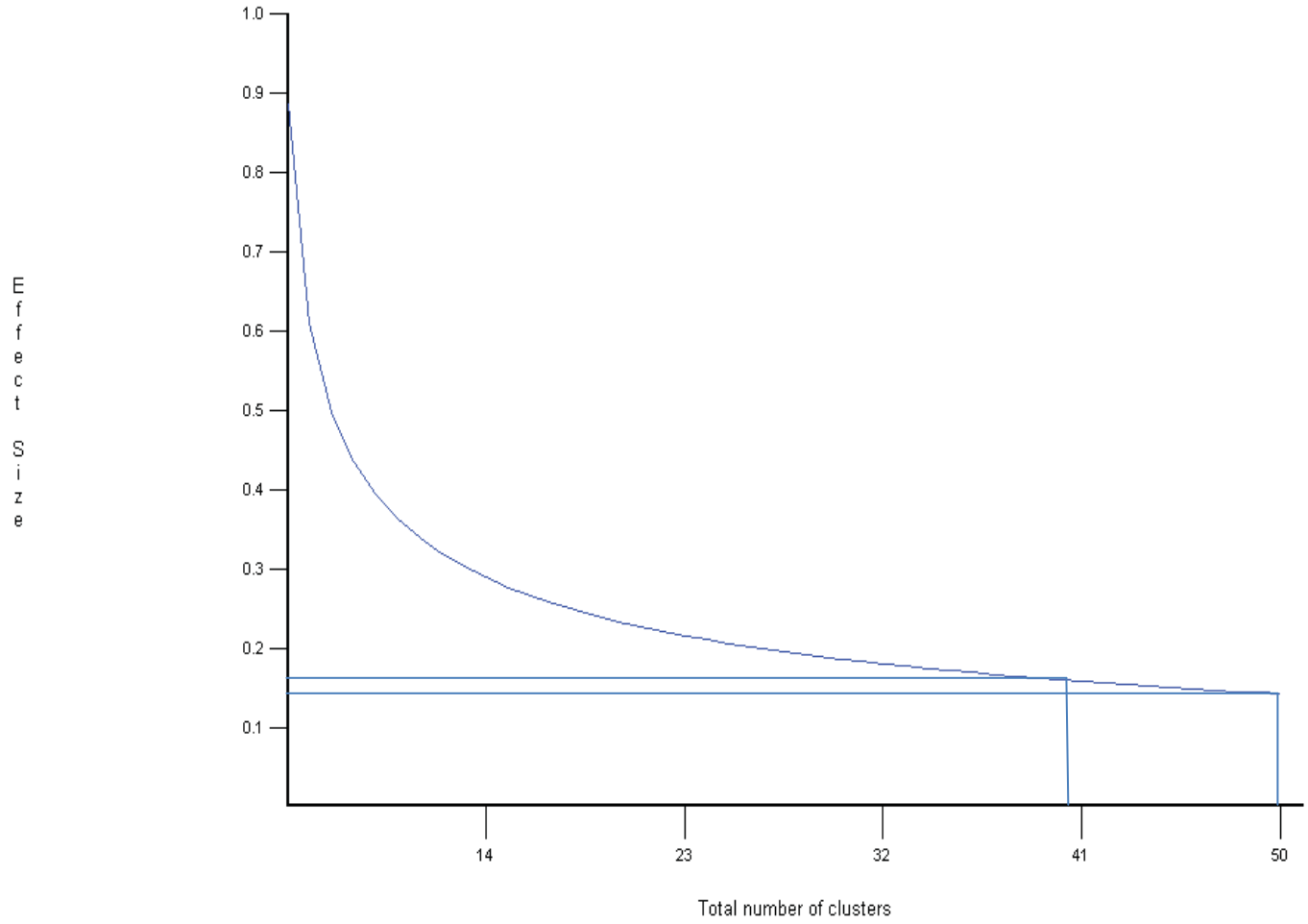
Effect Size
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
14
23
32
41
50
Total number of clusters

Figure A.16.b

Figure A.17: *Minimum detectable effect size calculation for continuous outcome variable overall perceived control.*

From empirical analyses: ICC = 0.027

Assuming power=0.8, proportion of the variance explained by group level covariates: 0.4. Number of subjects in each platoon is 28 (participation rate=50%, Figure A.17.a) to 39 (participation rate=70%, Figure A.17.b). Significant level: a=0.05.

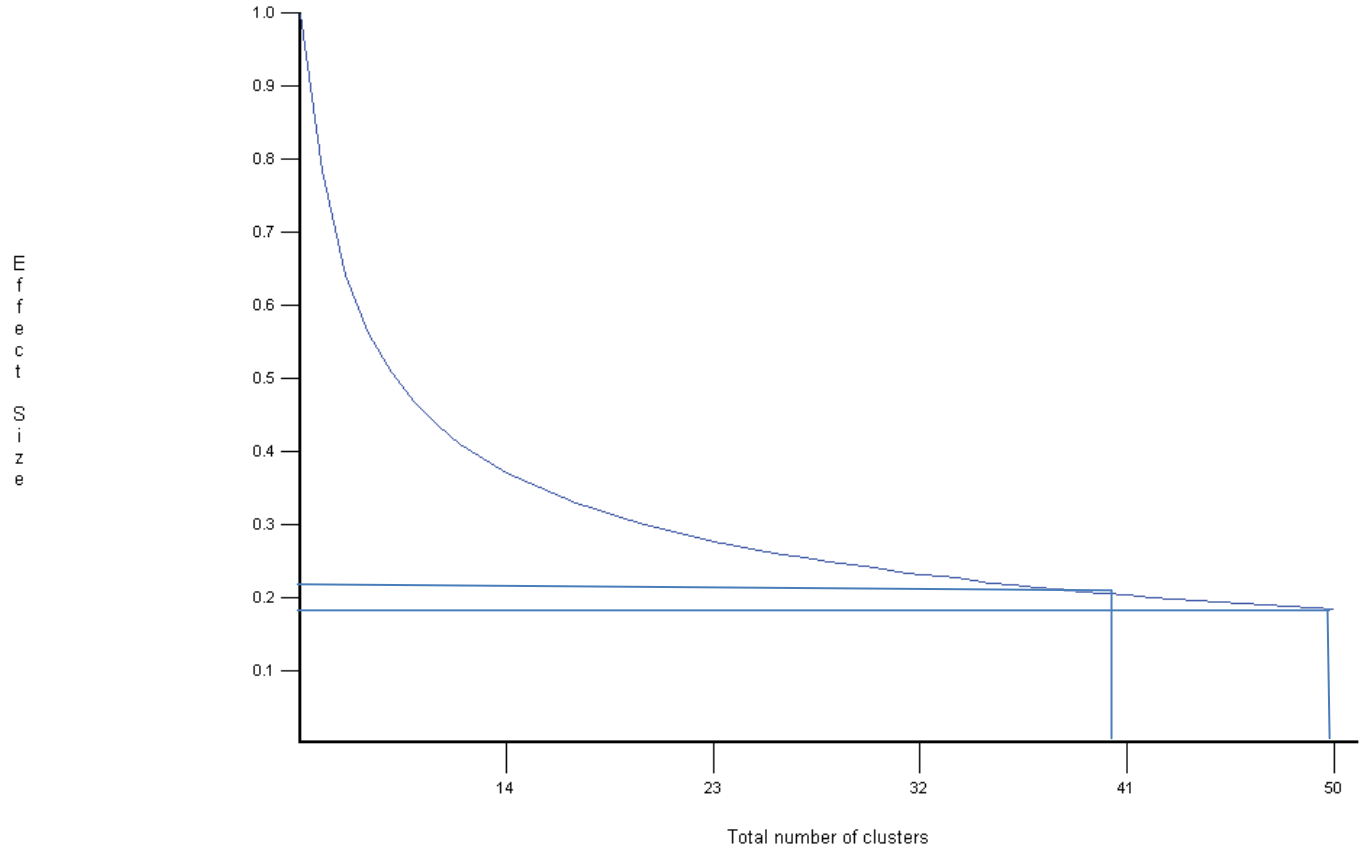

Figure A.17.a

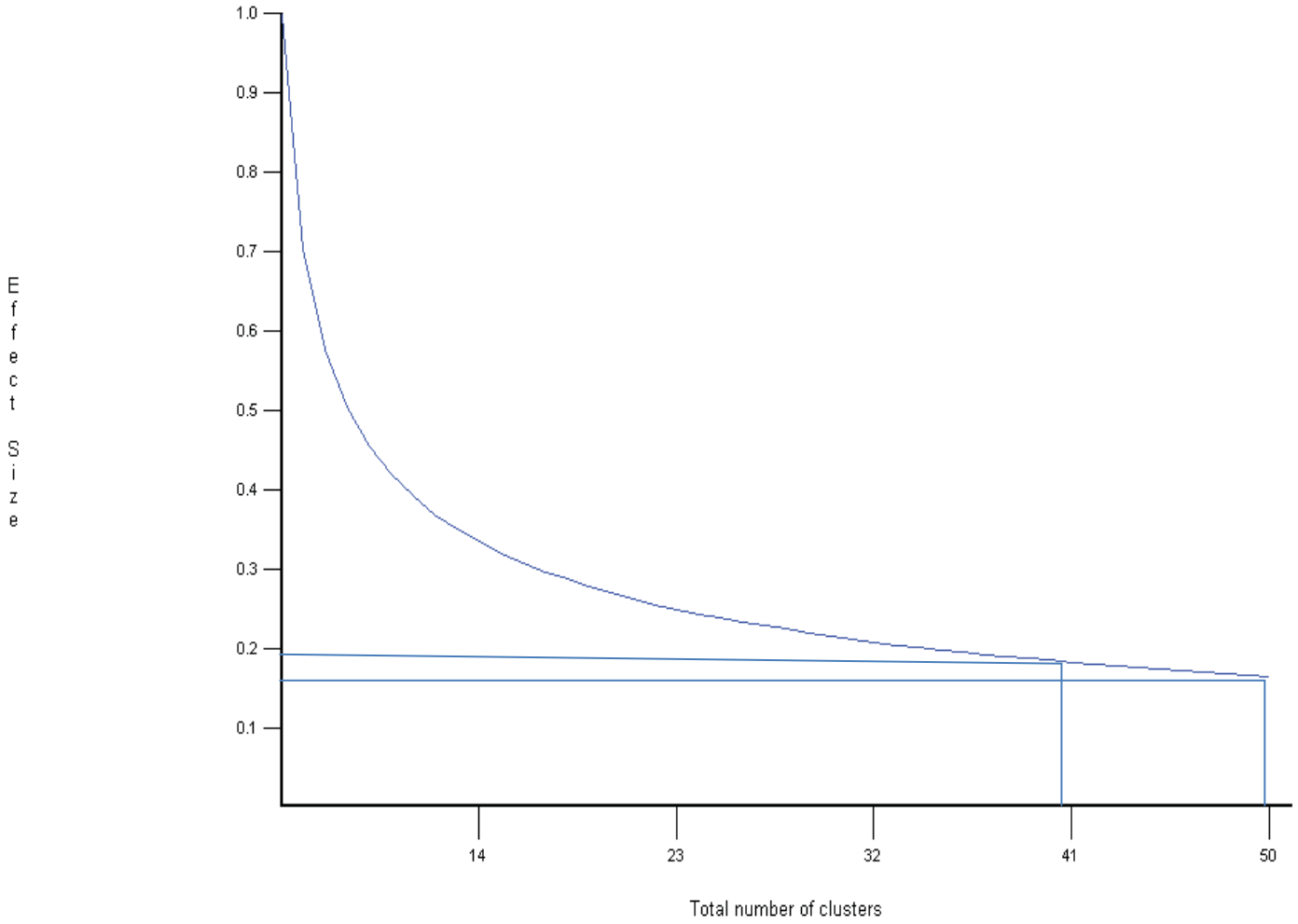
Effect Size
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
14
23
32
41
50
Total number of clusters

Figure A.17.b

Figure A.18: *Minimum detectable effect size calculation for continuous outcome variable perceived control.*

From empirical analyses: ICC = 0

Assuming power=0.8, proportion of the variance explained by group level covariates: 0.4. Number of subjects in each platoon is 28 (participation rate=50%, Figure A.18.a) to 39 (participation rate=70%, Figure A.18.b). Significant level: a=0.05.

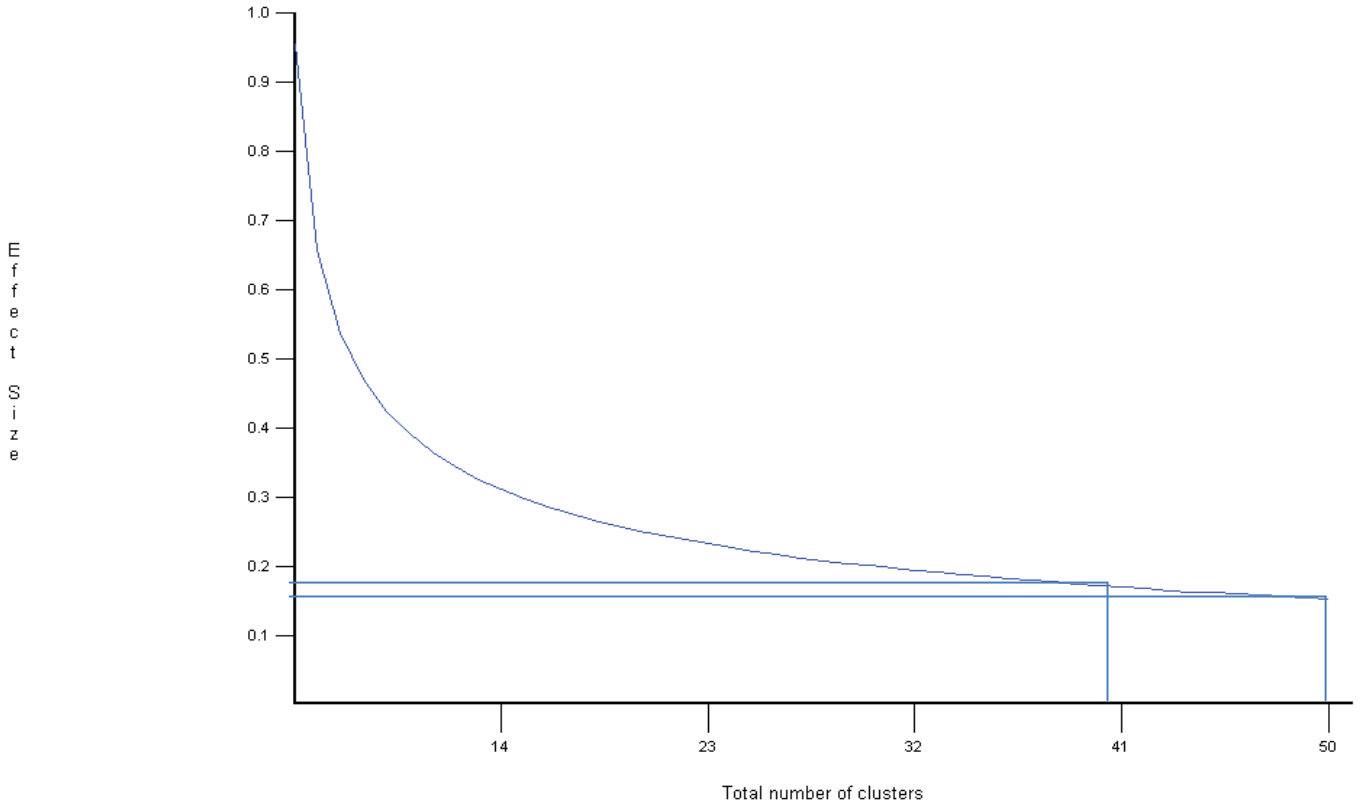

Figure A.18.a

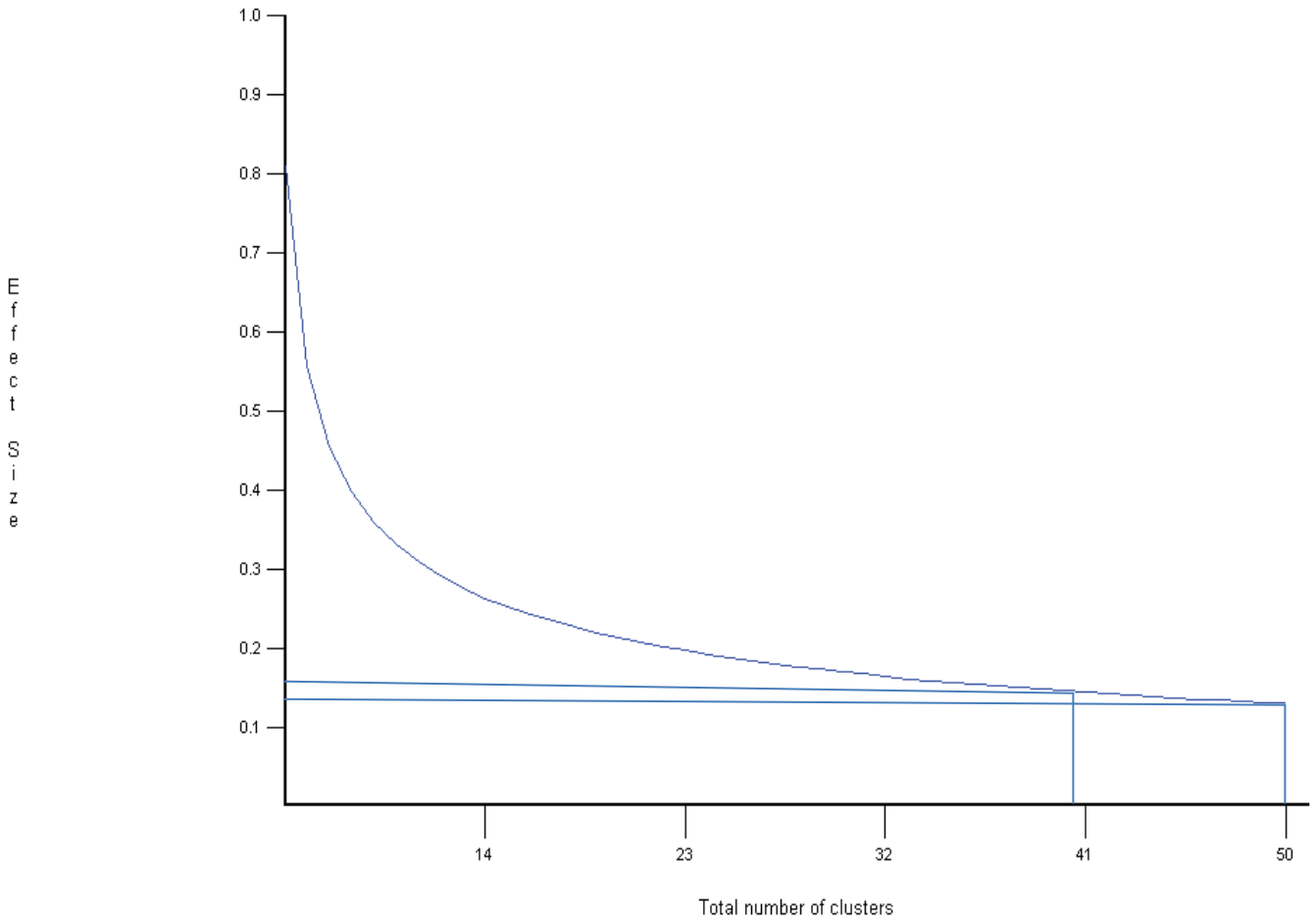
Effect Size
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
14
23
32
41
50
Total number of clusters

Figure A.18.b

Figure A.19: *Minimum detectable effect size calculation for continuous outcome variable self-efficacy.*

From empirical analyses: ICC = 0.038

Assuming power=0.8, proportion of the variance explained by group level covariates: 0.4. Number of subjects in each platoon is 28 (participation rate=50%, Figure A.19.a) to 39 (participation rate=70%, Figure A.19.b). Significant level: a=0.05.

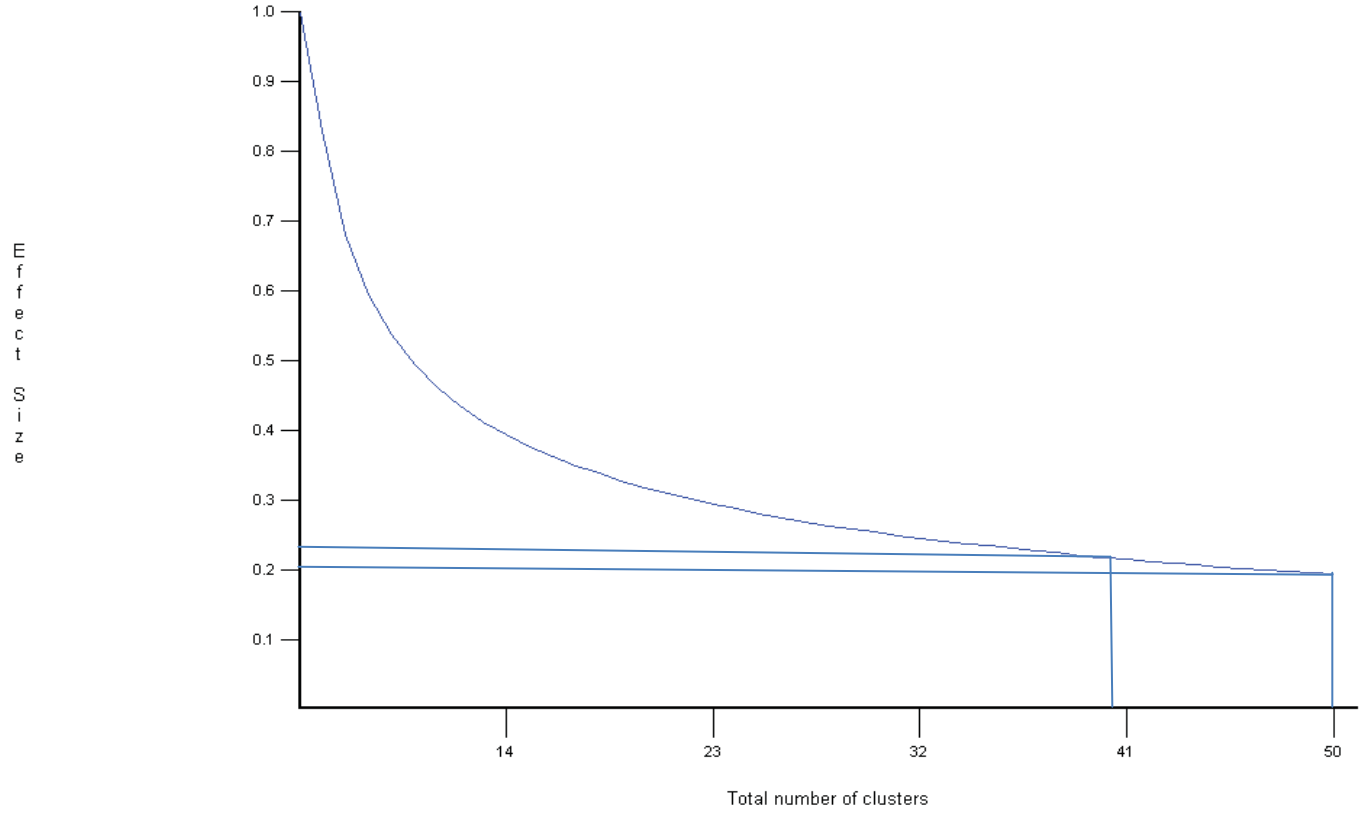

Figure A.19.a

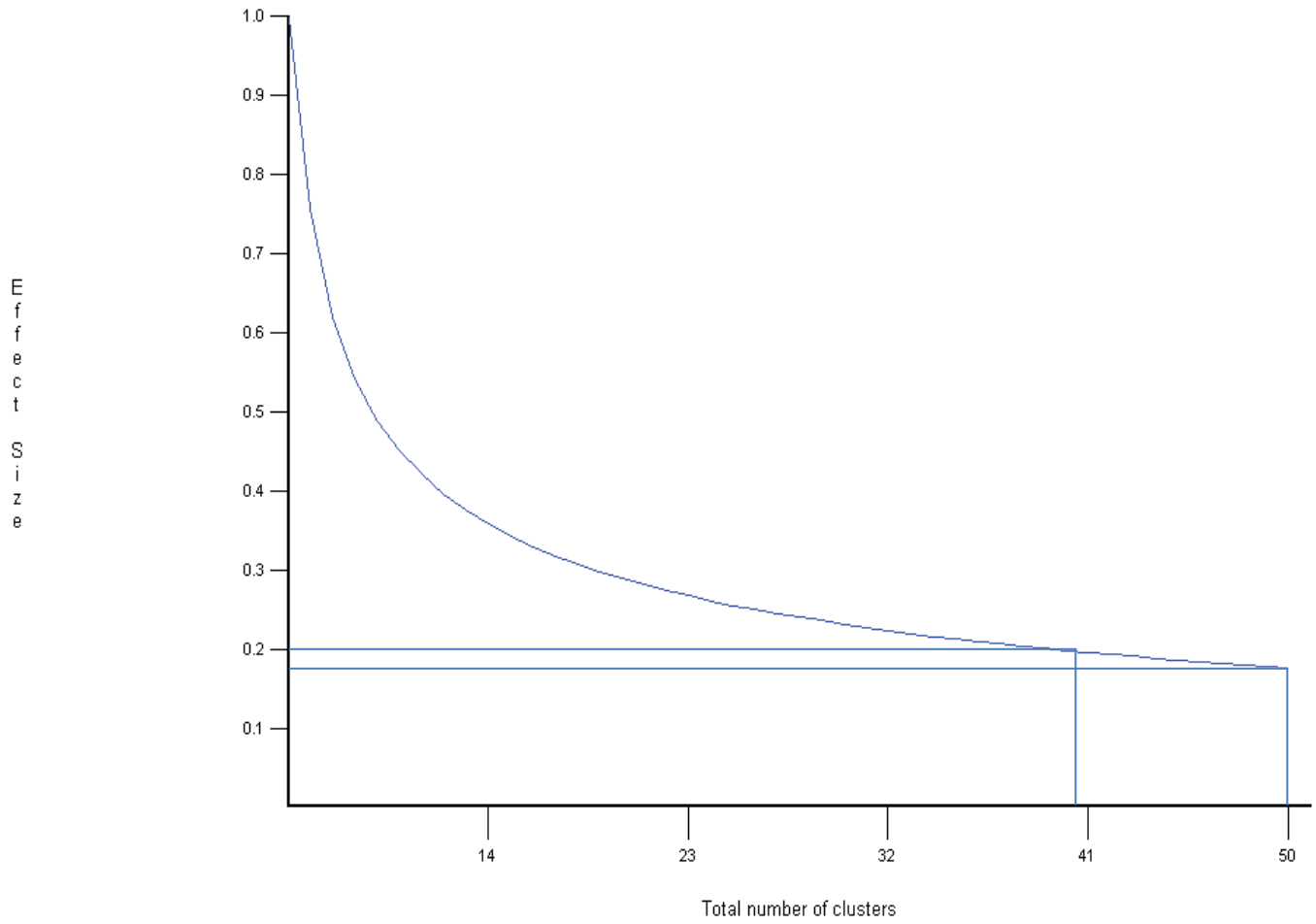
Effect Size
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
14
23
32
41
50
Total number of clusters

Figure A.19.b

Figure A.20: *Minimum detectable effect size calculation for continuous outcome variable PHQ9.*

From empirical analyses: ICC = 0.007

Assuming power=0.8, proportion of the variance explained by group level covariates: 0.4. Number of subjects in each platoon is 28 (participation rate=50%, Figure A.20.a) to 39 (participation rate=70%, Figure A.20.b). Significant level: a=0.05.

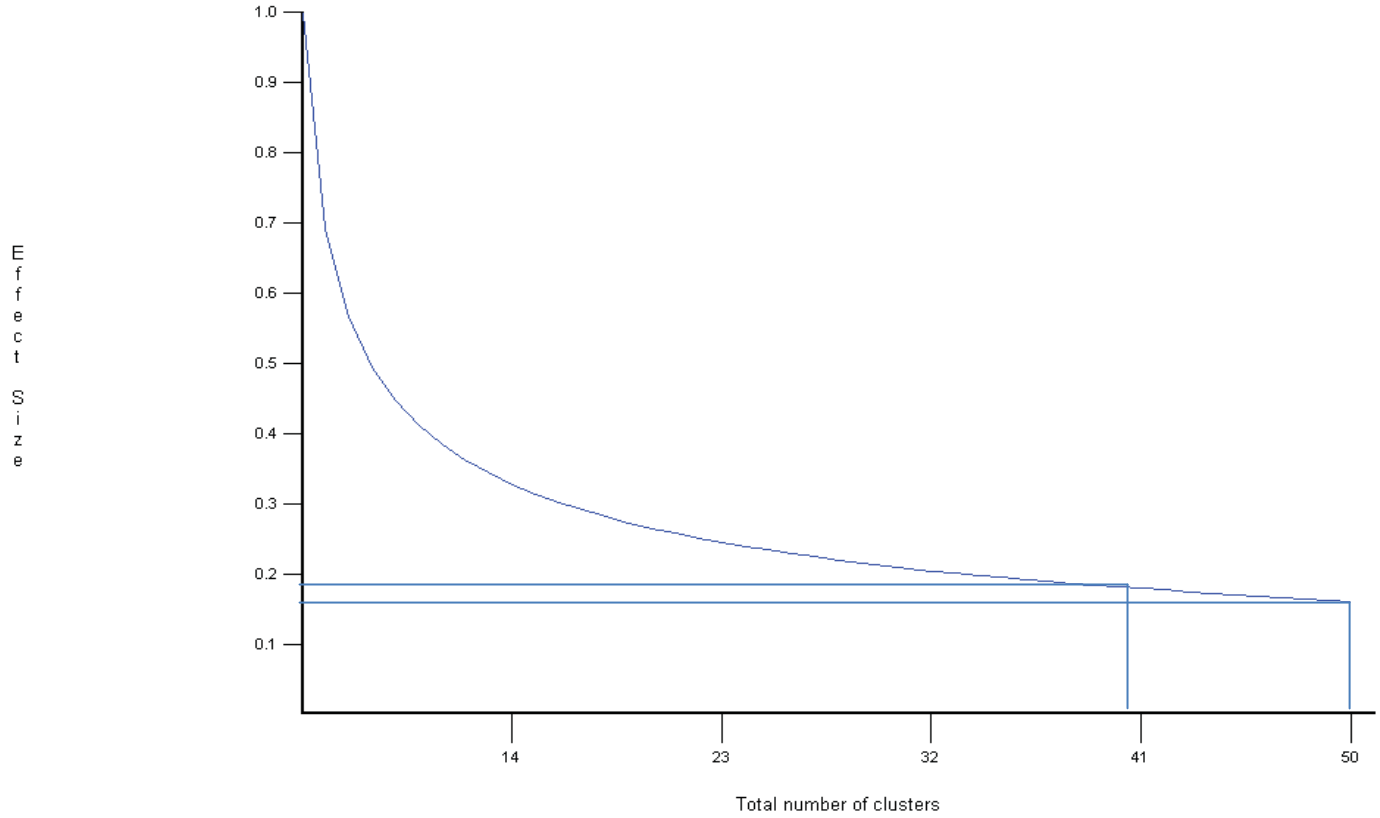

Figure A.20.a

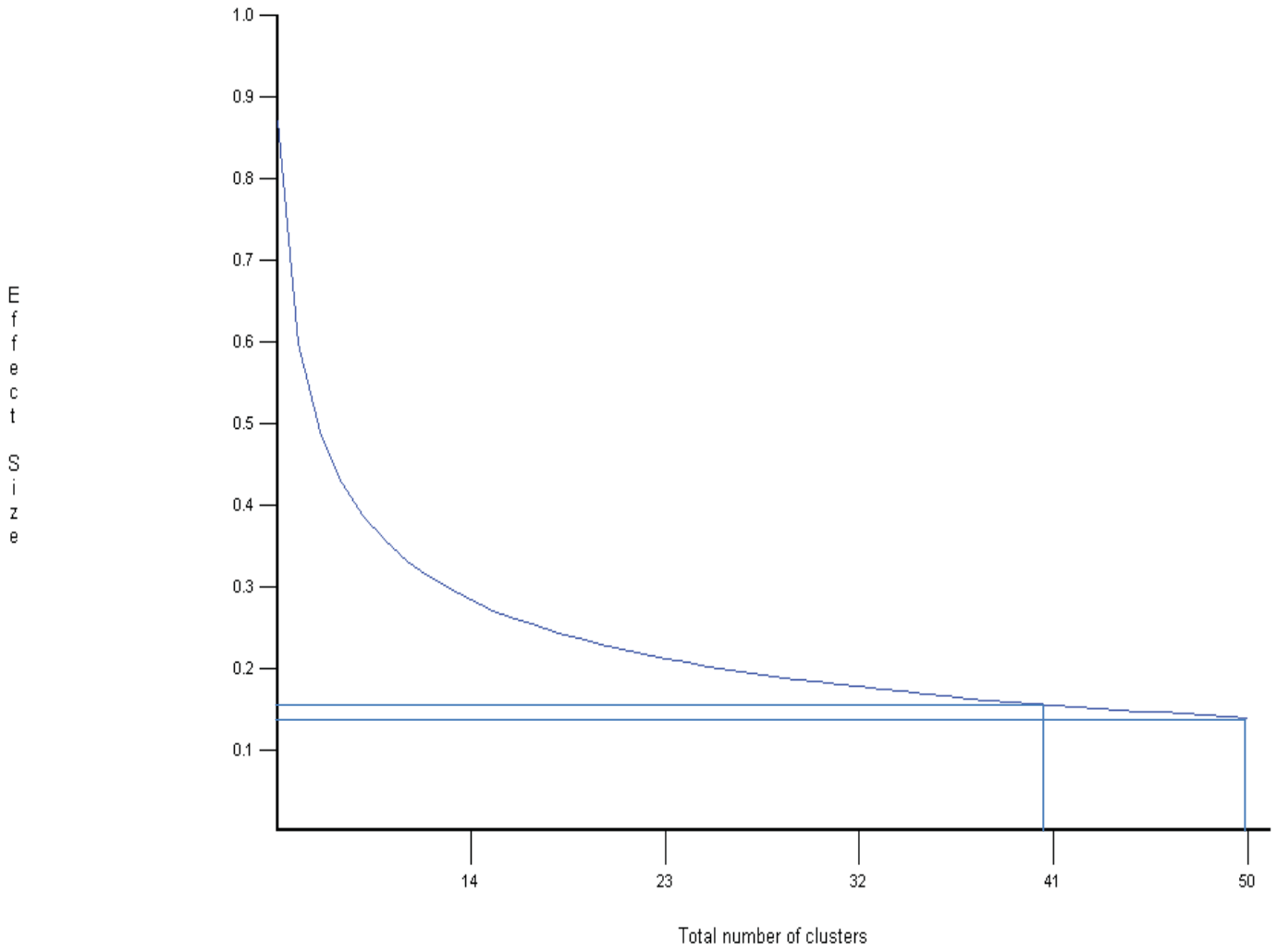
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
Effect Size
14
23
32
41
50
Total number of clusters

Figure A.20.b

Figure A.21: *Minimum detectable effect size calculation for continuous outcome variable PHQ15.*

From empirical analyses: ICC = 0.025

Assuming power=0.8, proportion of the variance explained by group level covariates: 0.4. Number of subjects in each platoon is 28 (participation rate=50%, Figure A.21.a) to 39 (participation rate=70%, Figure A.21.b). Significant level: a=0.05.

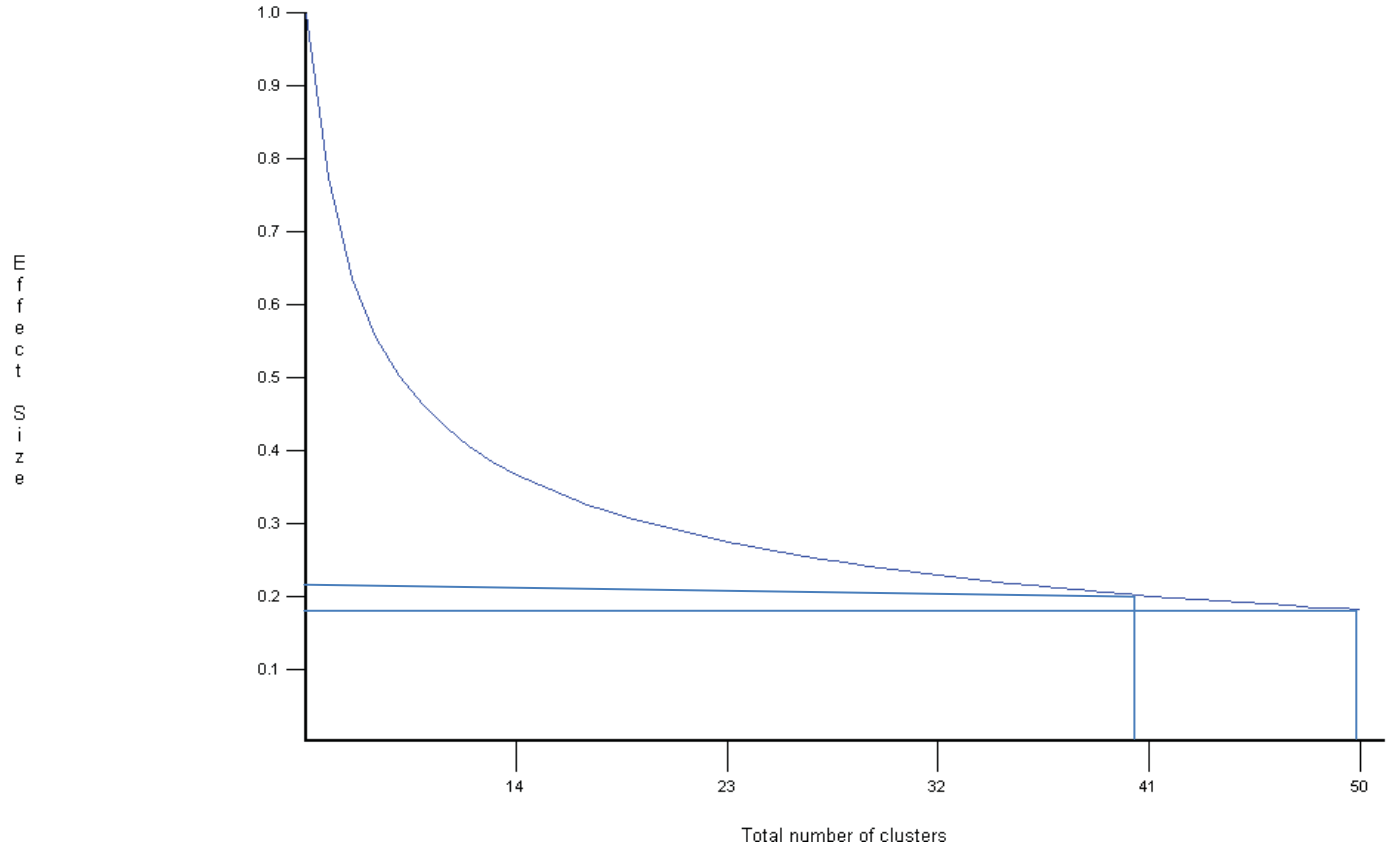

Figure A.21.a

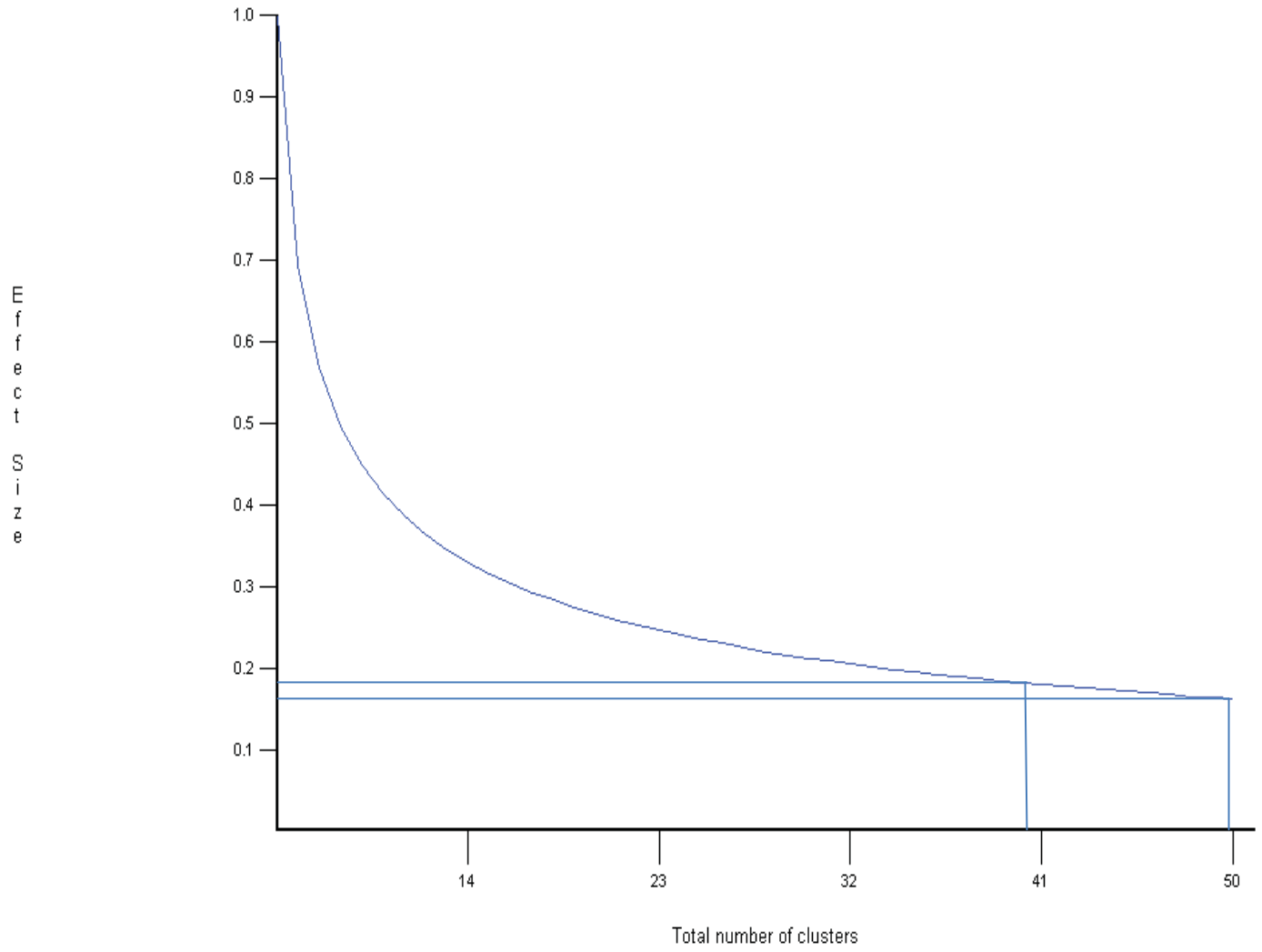
Effect Size
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
14
23
32
41
50
Total number of clusters

Figure A.21.b

Figure A.22: *Minimum detectable success rate among the intervention group calculation for binary outcome variable BMQ graduation.*

From empirical analyses: BMQ graduation rate = 80%

Assuming the success rate among control group is 0.80. 95% confidence interval for the success rate in control group is 0.75 – 0.85. Power=0.8. Number of subjects in each platoon is 21 (participation rate=50%, dropout rate=15%, Figure A.22.a) to 29 (participation rate=70%, dropout rate=15%, Figure A.22.b).

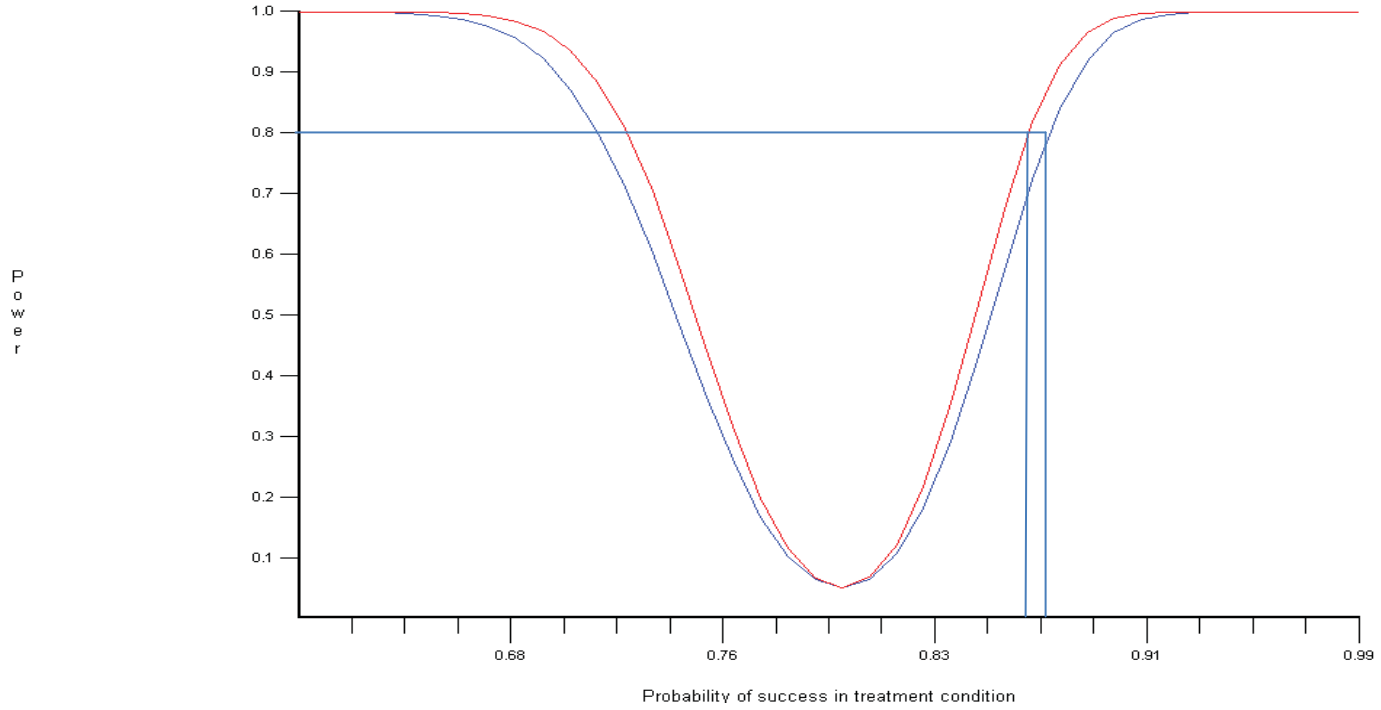

Figure A.22.a

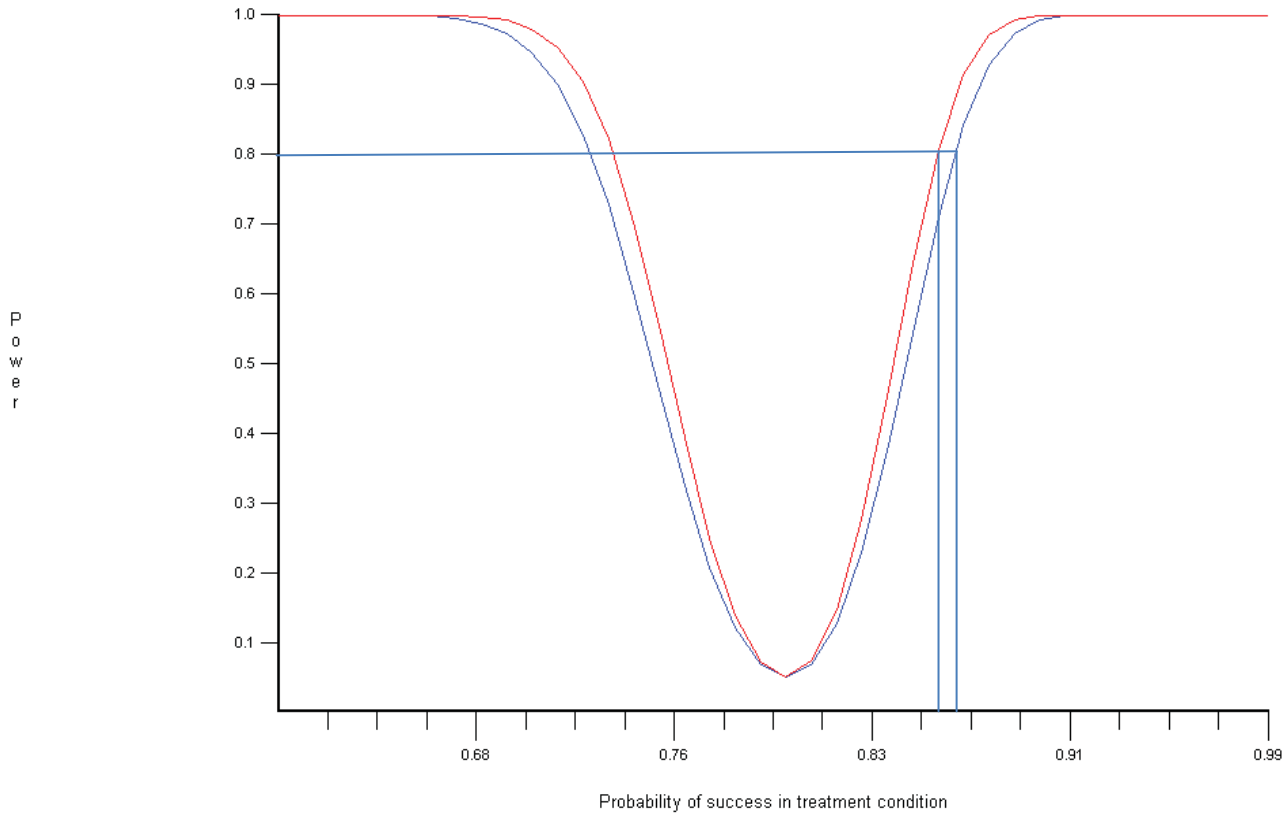
Power
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.68
0.76
0.83
0.91
0.99
Probability of success in treatment condition

Figure A.22.b

This page intentionally left blank.

List of symbols/abbreviations/acronyms/initialisms

BMQ	Basic Military Qualification
CAF	Canadian Armed Forces
CAF-MHSUQ	Canadian Armed Forces Recruit Mental Health Service Use Questionnaire
CFLRS	The Canadian Forces Leadership and Recruit School
DND	Department of National Defence
DRDC	Defence Research and Development Canada
DSTKIM	Director Science and Technology Knowledge and Information Management
GRCT	Grouped randomized controlled trial
ICC	Intra-class correlation coefficient
PHQ	The Patient Health Questionnaire
R2MR	Road to Mental Readiness

This page intentionally left blank.

DOCUMENT CONTROL DATA

(Security markings for the title, abstract and indexing annotation must be entered when the document is Classified or Designated)

1. ORIGINATOR (The name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g., Centre sponsoring a contractor's report, or tasking agency, are entered in Section 8.)

Defence Research and Development Canada – Toronto
1133 Sheppard Avenue West
P.O. Box 2000
Toronto, Ontario M3M 3B9

2a. SECURITY MARKING (Overall security marking of the document including special supplemental markings if applicable.)

UNCLASSIFIED

2b. CONTROLLED GOODS

(NON-CONTROLLED GOODS)
DMC A
REVIEW: GCEC DECEMBER 2012

3. TITLE (The complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S, C or U) in parentheses after the title.)

Power analysis for a proposed Group Randomized Control Trial (GRCT) on the Road to Mental Readiness (R2MP) program

4. AUTHORS (last name, followed by initials – ranks, titles, etc., not to be used)

Liu, A.; Fikretoglu, D.

5. DATE OF PUBLICATION (Month and year of publication of document.)	6a. NO. OF PAGES (Total containing information, including Annexes, Appendices, etc.)	6b. NO. OF REFS (Total cited in document.)
August 2014	75	28

7. DESCRIPTIVE NOTES (The category of the document, e.g., technical report, technical note or memorandum. If appropriate, enter the type of report, e.g., interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Scientific Report

8. SPONSORING ACTIVITY (The name of the department project office or laboratory sponsoring the research and development – include address.)

Defence Research and Development Canada – Toronto
1133 Sheppard Avenue West
P.O. Box 2000
Toronto, Ontario M3M 3B9

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)	9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.)

10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document.)	10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)
DRDC-RDDC-2014-R68	

11. DOCUMENT AVAILABILITY (Any limitations on further dissemination of the document, other than those imposed by security classification.)

Unlimited

12. DOCUMENT ANNOUNCEMENT (Any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in (11) is possible, a wider announcement audience may be selected.))

Unlimited

13. ABSTRACT (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

The Road to Mental Readiness (R2MR) program is the largest mental health training initiative in the Canadian Armed Forces (CAF). As part of an effort to test the efficacy of R2MR at Basic Military Qualification (BMQ) with a group randomized control trial (GRCT), we conducted a robust power analysis to determine the sample size that would be required for the GRCT on R2MR. We also calculated intraclass correlation coefficients (ICCs) for the outcomes that will be measured in the GRCT, a necessary preliminary step for the power analysis. Data from the calculation of the ICCs were extracted from multiple programs of ongoing research with the Non-Commissioned Member (NCM) recruits, the intended target population for the GRCT. The results of our analyses suggest that data collected over the course of one full fiscal year will yield sufficient statistical power to detect expected effect sizes for most but not all of our outcomes. We therefore recommend data collection lasting up to one and a half years for the proposed GRCT on R2MR.

---

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g., Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

mental health training, R2MR, efficacy